Vol. 40, No. 5 October, 1950

Journal of Experimental Psychology

EDITED BY

FRANCIS W. IRWIN

UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CLARENCE H. GRAHAM, *Columbia University*
HARRY HELSON, *Brooklyn College*
DONALD B. LINDSLEY, *Northwestern University*
DONALD G. MARQUIS, *University of Michigan*
GEORGE A. MILLER, *Harvard University*
L. L. THURSTONE, *University of Chicago*
G. R. WENDT, *University of Rochester*
JOSEPH ZUBIN, *N. Y. Psychiatric Institute, Columbia University*

CONTENTS

A Study of Simple Learning under Irrelevant Motivational-Reward Conditions: K. W. SPENCE, G. BERGMANN, AND R. LIPPITT 539

An Experimental Investigation of Transposition as a Function of the Difference between Training and Test Stimuli: T. S. KENDLER 552

The Effect upon Generalized Drive Strength of Emotionality as Inferred from the Level of Consummatory Response: A. AMSEL AND I. MALTZMAN 563

Response Latency at Zero Drive after Varying Numbers of Reinforcements: D. ZEAMAN AND BETTY J. HOUSE 570

The Effect of Remote Anchoring Points upon the Judgment of Lifted Weights: R. K. HEINTZ ... 584

Proactive Inhibition as a Function of Response Similarity: R. L. MORGAN AND B. J. UNDERWOOD 592

Reactive Inhibition as a Function of Number of Response Evocations: P. S. SIEGEL 604

The Influence of the Inter-Trial Interval on the Humphreys' 'Random Reinforcement' Effect during the Extinction of a Verbal Response: D. A. GRANT, J. P. HORNSETH, AND H. W. HAKE 609

The Relationship between Length of Interval Separating Two Learning Tasks and Performance on the Second Task: C. E. HAMILTON 613

Absence of Reminiscence in the Serial Rote Learning of Adjectives: C. E. NOBLE 622

Transfer of Verbal Training to a Motor Task: KATHERINE E. BAKER AND RUTH C. WYLIE 632

Shifting of Distribution of Practice in Mirror Drawing: J. C .TSAO 639

The Relationship between Order and Frequency of Occurrence of Restricted Associative Responses: W. A. BOUSFIELD AND W. D. BARCLAY 643

Retention of Pursuit Rotor Skill after One Year: H. M. BELL 648

Some Effects of Shape on Apparent Brightness: R. M. HANES 650

Accuracy of Visual Interpolation between Scale Markers as a Function of the Number Assigned to the Scale Interval: A. CHAPANIS AND M. LEVZOREK 655

The Visual Discrimination of Geometric Forms: R. C. CASPERSON 668

Figure-Ground Reversal as a Function of Visual Satiation: J. E. HOCHBERG 682

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STS., LANCASTER, PA.

AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY will, in 1951, have two volumes per year and monthly publication. The subscription rate per volume will be $7.00, or $14.00 annually. Single copies will be $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for psychological theory. Articles of a length exceeding 25 printed pages cannot be accepted.

With rare exceptions, articles are published in the order of their receipt. Authors are supplied with 50 free offprints without covers. All of the cost of author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor-elect, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

This JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, *including tables, footnotes and references.* Tables, figures, and footnotes must be placed on separate pages at the end of the manuscript.

Journal of
Experimental Psychology

VOL. 40, NO. 5 OCTOBER, 1950

A STUDY OF SIMPLE LEARNING UNDER IRRELEVANT MOTIVATIONAL-REWARD CONDITIONS

BY K. W. SPENCE, G. BERGMANN, AND R. LIPPITT

From the Psychological Laboratory, University of Iowa

INTRODUCTION

In the first report (22) of a series of experimental studies concerned with the motivational-reward conditions underlying learning in animals, the sign-gestalt (non-reinforcement) theory of trial-and-error learning as formulated by Tolman (25) and others (12, 29) was subjected to experimental test. According to this theory such selective learning involves the acquisition of sign-gestalt-expectations with respect to the various aspects (signs) of the situation, the alternative responses and the different significates to which each of the latter leads. These expectation-sets (also termed cognitions, hypotheses, etc.) constitute one of Tolman's so-called intervening or theoretical variables which are assumed to be determinants of behavior. Thus, in the alternative choice situation typical of simple trial-and-error learning, the response on any trial will depend upon the state of these cognitions (C) and another intervening variable, motivational state (M) or, more briefly, $B = f(C, M)$.

Cognition (C) and motivation (M) are, in Tolman's system, hypothetical state variables of the organism which are to be specified in terms of certain past experimental operations or conditions. The particular point at issue in our experiment was Tolman's hypothesis as to the conditions essential to the development of the cognitive states. Tolman and his group have taken the position that this learning factor (i.e., cognition) develops as the result of mere association by contiguity (sequence in experience) of the sign, the response and the significate, and that such learning is not due, as the opposing reinforcement theorists have claimed, to any rewarding property possessed by a significate. That is, according to this view, the learning of sign-significate expectations is not dependent upon their being a reward or goal object for which the organism is at the moment motivated. Any differential reward values among the significates will, Tolman admits, make for a difference in the response of the organism at any moment, but they play little or no part in the learning *per se*—i.e., the formation of the expectation-sets.

As a test of this theoretical position 10 white rats were run in a simple two-choice Y-maze, one alley (left) of which led to a

goal box containing food and the other (right) to a goal box containing a water bottle. The *S*s, motivated for water and satiated for food, were given training for 12 days, in which each *S* was given five trials per day. In two of the five trials the *S*s were free to choose either path and the remaining three trials were forced, one to the water and two to the food side.

On the day following the completion of the learning series, a critical test was given in which the motivation of the *S*s was shifted from thirst to hunger. According to Tolman's hypothesis, the *S*s should have learned what object (food or water) each of the two alleys led to and when motivated for food they should have been expected to reveal this learning by taking the alley leading to the food. The results of the test, however, were entirely negative to this theoretical implication, for every animal, although hungry, chose the water alley. Futhermore, it was found that this group of subjects subsequently were unable to learn to go to the food alley more readily than a comparison group which had had the same previous training experience of 12 days, except that no food had ever been present in the left goal box. Both of these results are in line, it may be noted, with the implications of the reinforcement type of theory such as that of Hull (6) and Thorndike.

A second study in this series essentially similar in design to this investigation has been carried out by Kendler (9). The maze employed in this instance was of the simple T-type and further differentiation of the critical signs was achieved by having one of the alleys painted black while the other remained unpainted. Six of the subjects were trained under thirst, as in the Spence and Lippitt study and four under hunger drive. In the test series involving a shift of motivation all subjects continued to run to the side which had led to the appropriate reward object during the training series.[1]

[1] The results for Kendler's subjects that were trained under hunger in the learning series and then tested under thirst effectively refute the recent criticism of Hayes (2). The latter claimed that the subjects of Spence and Lippitt had had no experience with a strong hunger drive prior to the test and could not be expected to react appropriately toward food on the test trial. Hayes claims to have shown that rats trained originally on food did respond appropriately to water on the first test trial. He suggested, therefore, that the drive sequence training on hunger, testing on thirst would be adequate. The identical results of Kendler under both conditions refute Hayes' criticism. Furthermore, the results of the present experiment, as will be seen, seriously call into doubt Hayes' claim that rats will not be able to respond appropriately when tested under hunger.

While the findings of these experiments were quite contrary to what Tolman's theory would lead one to expect, it was suggested that certain modifications might be made in his formulation that would permit one to maintain a non-reinforcement position and yet account for the present experimental findings satisfactorily. One possibility that suggested itself was that the learning involved, in addition to, or possibly even instead of, the sign-significate expectations with respect to food and water, the formation of the expectations that the right-side alley led to a need-satisfying situation without differentiation as to its specific nature, whereas the left alley led to the expectation of failure of need satisfaction. Such 'knowledge,' no matter what the motivational state, would presumably lead to the taking of the alley on the right during the test trial.

Statement of Problem

The experiment described in the present report attempted to eliminate the differential need-satisfying character of the two end boxes that existed in our first study. One obvious way of accomplishing this was to have the *S*s satiated for both food and water during the training runs. A serious difficulty with such a procedure is readily apparent, however, in that one

does not have any drive state that will motivate the animal to make the necessary responses in the situation. In the hope that some mild form of motivation (possibly an exploratory drive) might operate long enough to make the animals enter the alleys and end boxes a sufficient number of trials during the training period, a group of *S*s was started under this condition of satiation for both food and water. Furthermore, in order not to have the end boxes serve as a possible goal-situation for the animals by reason of the fact that entrance into them led to their being picked up and placed in their home cage or a cage with other animals (social goal), the procedure was followed of placing each *S*, on removal from the goal box, in the same type of small individual carrying cage that it was in prior to the start of the run. If such a situation would lead to the necessary behavior taking place, it was felt that it would provide the ideal experimental conditions, for the necessary sequence of events (S_L—R_L—S_F, S_R—R_R—S_W) would have occurred without presumably any goal reinforcement at the end boxes.

Unfortunately, the exploratory drive, or whatever the motivation operating at first was, did not persist sufficiently long to provide for an adequate number of experiences in the maze alleys. After making two or three 'exploratory' runs into the alleys the *S*s refused to run in the maze but instead would lie down in the starting alley and apparently go to sleep. Some attempt was made to keep the *S*s moving by gently pushing them, but this was abandoned as unsuccessful.

With this failure to obtain satisfactory runs in the training condition without the use of some goal situation, it was decided to introduce a goal situation that would be identical (equalized) for both end boxes. The goal situation hit upon was a cage similar to the home cage of the animals, in which all except the *S* run first on the day found one or more of his home-cage mates. As the order of running the *S*s was changed each day, all animals found themselves being placed in this 'social' situation after reaching one or other of the end boxes on most of their trials. This procedure apparently provided for some type of motivation and goal reward, for under it the subjects did *perform* much more satisfactorily than in the case of the first group of satiated subjects. On the occasions that an *S* would not run in a training trial, it was prodded until the run was completed. There were considerable individual differences in the extent to which the *S*s were motivated to get to the end box and eventually to the social cage, and on the whole the times taken to run were longer than in the first experiment in which the *S*s were motivated by thirst.

The fact that the conduct of the experiment required a motivating condition and a goal situation is of no little significance for the theoretical point at issue. For one thing, it meant that the arrangement did not provide for the sequence in experience of the sign, response, and significate *without the occurrence of a reward or goal situation following the sequence.* Thus, the question of whether the mere contiguity of these events in experience without a reinforcing state of affairs is sufficient for the establishment of the sign-gestalt-expectations with respect to food and water cannot be answered by this experimental design. Our experiment can only tell us whether expectation-sets concerning significates for which there is not at the moment any need, can be established on the basis of some other motivating state and accompanying significate which does have reward value. For example, is the white rat able to develop expectations with respect to the food and water while motivated only for the 'social' cage which it gets to upon entering either of the end

boxes? It is apparent, then, that while we have equalized the reward values of the two end boxes, we are not testing the question of whether sign-gestalt expectations can be established *independently* of motivation and reward.

Experimental Procedure

Subjects.—The *S*s, 39 white rats from the departmental colony, consisted of 15 males and 24 females. Ages at the beginning of the preliminary training varied slightly around 90 days.

Apparatus.—A ground plan and detailed description of the apparatus was given in the earlier article (22). It is sufficient to say here that it was a single-choice Y-maze with a relatively large square area at the choice point and a jump of two in. at the entrance to each of the alleys. One of the goal boxes was fitted with a water bottle similar in design and appearance to the type used in the living cages of the animals. The food goal box contained a sliding food tray in the floor but this was more or less completely covered by some six to eight large pellets of Purina Dog Chow spread over the floor.

Vertically sliding doors, operated from behind a one-way screen, prevented the *S*s from seeing the objects in the goal boxes and from retracing in the maze. Differentiation of the two alleys from one another was increased by having the floor at the entrance to the left alley covered with fine hardware cloth while the entrance to the right alley was covered with a very much coarser screening.

Preliminary training.—After several days of handling, the *S*s were given an opportunity to explore and become familiar with the maze. During this experience they were satiated for both food and water. Six to ten *S*s were placed in the maze at the same time and were allowed to explore (1) the starting box, choice chamber and one alley, (2) the second alley, and (3) both alleys simultaneously for about 30 to 40 min., respectively. During this period the water bottle was removed and the end of the box containing it covered over. No food was in the food goal box.

On the following day, each *S* was given two runs alone in the apparatus in order to ascertain information as to a position preference. On the basis of the choices made, each *S* was classified as a left, right, or neutral animal. The maze was in the same condition as day 1 and the *S*s were satiated.

Satiation routines.—The following feeding and drinking schedule was begun two days prior to the first day of preliminary training, and was used throughout the training trials. About nine hours prior to the time of running, which began between 4:30 to 5:00 P.M. each day, the water bottles were taken from the living cages. Approximately three hours before the time of running, the water bottles were placed back on the cages and an additional water dish was put into the cage. An abundance of food was also placed in the cage. About one hour before running time an extra dish of small pellets (Purina Dog Chow) was put in the cage and the water dishes refilled. Following completion of the training runs, water was available until the next morning, but no food was given until the following day.

The satiation procedure during the two test trials following the completion of training was as follows. Prior to being tested under hunger motivation, food was withheld for 23 hours. The water bottle was available at all times in the cage and the procedure of placing an additional pan of water in the living cage before the daily run was used. In the case of thirst motivation, the water bottle were removed approximately nine hours prior to the test run. Food was available in the cage at all times and small pellets were placed in the living cage before the test trial.

Learning series.—The *S*s were given four trials daily for seven days. In the first and third trial of each day the *S*s were free to choose either alley. On the second and fourth trial they were forced to take the alley not chosen on the previous trial. An equal number of experiences in each alley was thus provided all subjects.

The animals were run in groups varying in size from six to ten *S*s. The same order of running was followed on successive trials of a day but differed from day to day. All *S*s of a group were run before the next trial was given, a procedure which provided for a minimum interval of at least 10 min. between the successive trials for each *S*.

As was described in the section on the statement of the problem, the motivation for running the maze was the 'need' to get to the social cage containing cage mates. The *S*s were taken to the experimental room in a carrying cage which consisted of a number of small individual compartments (4 in. by 10 in.). *S* was taken out of this carrying cage and placed directly in the starting box of the maze. At the completion of the trial it was placed in the social cage. In the case of trials two, three and four, each *S* was taken out of the social cage three runs before its turn to run and placed in one of the narrow compartments of the carrying cage. After the last trial of the day, a five-min. period elapsed before the group as a whole was returned to its living cages.

Upon entering an end box *S* was permitted to explore until *E* was sure that it had examined the food or the water orifice. As the food was spread out over the floor of the goal box, the *S*s were forced to walk over it and there was little difficulty in determining whether they sniffed at it. In the case of the water spout, however, *S* was not so likely to come upon it in the course of its exploration. If at the end of a 15-sec. period the water spout had not been approached, the *S* was picked up and placed directly before the spout with its nose just a fraction of an inch from it. Typically the *S*s moved around the goal box sniffing at either the food or water spout within the 15-sec. period. Care was taken to note whether the animal ate any of the food or drank water. No *S*s included in the 39 cases were ever observed to eat or drink.

Test series.—Test trials were given on the two days following the training period. On these trials the *S*s were motivated on one day for one of the goal objects and satiated for the other. On the following day the motivation was reversed. If the *S* chose correctly (i.e., chose the alley which led to the goal object for which motivated) on the first trial of the day, it was permitted, depending on its motivation, either to drink for five to ten sec. or to eat for the same period. It was then placed in the social cage and after a 10-min. period returned to the home cage. If the first test trial was incorrect, the *S* was given further trials until the correct alley was taken. The same procedure was followed on the second test given on the following day.

In the first four groups of subjects run, the decision as to which motivation was to be present in a particular *S* on the first trial was determined by the preference for the two alleys shown in the fourteen free choices of the training series. In all instances, if an *S* had chosen one of the alleys less than 50 percent (six times or less), it was motivated for the goal object in that alley. In the last group of ten subjects the reverse procedure was employed. In this group the *S*s were motivated for the goal object to be found in the alley for which a preference had been shown —i.e., eight or more choices of it made. In the case of five *S*s which chose each alley 50 percent of the time on the training trials, three were arbitrarily made thirsty and two hungry.

Results

Fig. 1 presents the results for all 39 *S*s in terms of the percentage of responses made on the first trial of each training day and the two test days to the side containing the goal object appropriate for *the motivation of the first test.* These values tend to run below

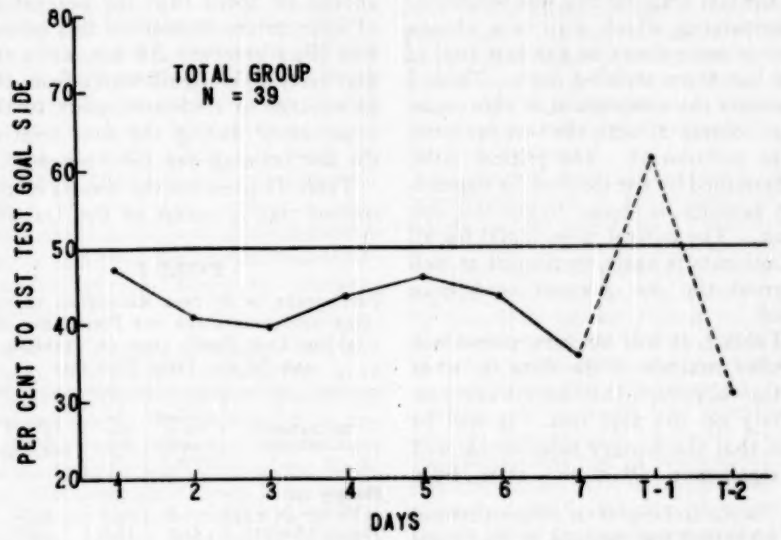


Fig. 1. Percentages of *S*s that chose the side appropriate on test I during seven training days and two test days

the 50 percent line for the training days because, as was indicated previously, the majority of *Ss* were motivated on test one for the goal object on their non-preferred side.

The marked shift in the curve at the first test day suggests that the subjects had learned something about the location of the two goal objects during this training period and when motivated for one of them, tended to take the alley leading to it. Twenty-four (61.5 percent) of the *Ss* chose the alley leading to the appropriate goal on this first test, as compared with only 33.3 percent choice of this alley on the first trial of the last day of training. The difference of 28.5 percent between training day seven and the first test day is 2.85 times its standard error, a critical ratio which is significant beyond the one percent level of confidence.[2]

As a further check, a more stable measure of the training preference than that provided by the first trial of the last training day was sought by ascertaining which arm was chosen two or more times on the first trial of the last three training days. Table I presents the comparison of this measure (column 2) with the test performance (column 3). The critical ratio, determined by the method for dependent samples, is shown in the last column. The critical ratio (2.98) for all 39 subjects is again significant at well beyond the one percent confidence level.

Table I, it will be seen, presents a further analysis of the data in terms of the subgroups that were hungry or thirsty on the first test. It will be seen that the hungry subjects showed a significant ($P < .01$) shift from their training performance on test 1, whereas the shift for the thirsty subgroup was not significant ($P > .14$). The poorer result for the latter group is not unlikely due to the fact that a thirst drive based on only nine hours of water deprivation was used in this experiment. This degree of thirst was selected on the basis of the obstruction box studies as being approximately equal in strength to a 24-hour hunger drive, but these results suggest that this may not be the case.

Looking at the results for the second test, in which the motivation was shifted from hunger to thirst, or vice versa, it will be observed in Fig. 1 that the choice point behavior of the subjects shifted markedly again. Only 30.8 percent of the *Ss* chose the alley leading to the goal object of test one, while 69.2 percent chose the alley leading to the appropriate goal for the second test. While this shift in choice of the two alleys was appropriate to the motivational change, it should be noted that the percentage of appropriate choices on this second test (69.2 percent) did not, as in the first test, differ significantly from the percentage of responses made to the same alley during the first trial of the last training day (66.7 percent).

Table II presents the results of the second test in terms of the training

TABLE I

PERCENTAGE OF *Ss* THAT RESPONDED TO THE SIDE APPROPRIATE FOR THE FIRST TEST ON (a) THE LAST THREE DAYS OF TRAINING, AND (b) THE FIRST TEST DAY

Motivational Conditions	(a) Last Three Training Days %	(b) Test Day %	Critical Ratio Diff./$\sigma_{Diff.}$
Hungry and Thirsty ($N=39$)	30.8	61.5	2.98
Thirsty ($N=23$)	34.7	56.6	1.46
Hungry ($N=16$)	25.0	68.7	2.64

[2] The standard error of the difference between the percentages was computed by the formula for dependent samples as described by McNemar (16, p. 77).

TABLE II

PERCENTAGE OF *Ss* THAT RESPONDED TO THE SIDE APPROPRIATE FOR THE SECOND TEST ON (a) THE LAST THREE DAYS OF TRAINING AND (b) THE SECOND TEST DAY

Motivational Conditions	(a) Last Three Training Days %	(b) Test Day %	Critical Ratio Diff./$\sigma_{Diff.}$
Hungry and Thirsty ($N=39$)	69.2	69.2	0
Thirsty ($N=16$)	75.0	68.7	−1.05
Hungry ($N=23$)	65.3	69.5	.32

performance measure based on the last three days of training. It will be observed that neither the group as a whole nor the two sub-groups shifted significantly from their training preference to the correct alley. Indeed, the *Ss* that were thirsty went fewer times to the appropriate side of the test day than they had gone on the last three days of training.

DISCUSSION OF RESULTS

Since the completion of this experiment in 1940 (**21**), a number of studies (**10, 15, 24**) essentially similar in principle and design to it have been carried out in the Iowa laboratory and one almost identical study has been reported by Meehl and MacCorquodale (**17**). We shall first present a summary of the findings of all these experiments and then discuss their implication for the rival theoretical interpretations.

The experiments of Kendler (**10**) and Meehl and MacCorquodale (**17**) gave almost identical results with those of the present study. Both obtained a significant number of appropriate responses on the first test, the results of the Meehl and MacCorquodale study being at the one percent level and that of Kendler at the five percent level of significance. Both studies also agreed with the present one in failing to obtain positive results on the second test in which the motivation of the first test was reversed. From the description of their experiment, the motivational-reward conditions of the Meehl and MacCorquodale investigation closely approximated those of the present study. The chief difference appears to have been that the *Ss* were put back in the home cage only after the second and fourth trials of the day. Whether they were placed in a cage containing other animals (social goal) after the first and third trials of the day is not clear. These experimenters explicitly recognize the operation of such reinforcements as "being removed from the confinement of the maze, returning to the home cage, etc." (**17**, p. 375).

Kendler's study, carried out in the Iowa laboratory, deliberately attempted to avoid the social goal box and substituted pushing of the *S* through the maze when it remained in one spot for more than five min. Presumably the motivation for locomotion in this study was the mild, noxious one resulting from the pushing. By proceeding to either end box the *S* escaped such prodding. It was then picked up from the maze and placed back in the carrying cage to await the next trial. The poorer results of this study might possibly be ascribed to the slower running of the *Ss* and hence the longer time elapsing between entering an alley and the subsequent experiencing of the food or water.

All three of these studies, then, gave positive evidence of the acquisition of sign-significate expectations with respect to the food and water when not motivated for these objects on the first test, but gave negative evidence on the second test. To complicate the picture even more, the two most recent studies (**15, 24**) in the Iowa laboratory have failed to give any favorable evidence whatever for such learning. Maltzman's (**15**) experiment attempted to improve the social reward conditions employed in the present experiment with a view to speeding up the running of the *Ss* and thus reducing the time between the occurrence of the cues at the entrance of an alley and the subsequent occurrence of the significates.

Other things equal, this presumably should provide better conditions for the acquisition of the essential cognitions. Maltzman did obtain such a speeding up of running at least over the present study and that of Kendler by placing the social cages, which contained a cage mate, at the end of each alley, one foot beyond the boxes containing the significates, food and water. The *S*s ran directly by means of a connecting section into the social cage. The latter contained a one-way door that prevented the social animal from entering the maze. Quite in contrast to the previous studies, however, Maltzman found that his *S*s did not respond appropriately to the side containing the significate for which they were motivated on the first test. Thus the correct side was chosen only by 17 out of 30 *S*s on the first test, a result obviously not significantly different from chance, which the *S*s had been running during the last five training days. The second test, involving a reversal of the motivation of test one, also gave results that did not differ significantly from chance, although 19 of the *S*s did respond appropriately. Two subsequent tests on the following two days also gave chance results.

So far as the motivational conditions are concerned, our most recent experiment (24) represents an extension of the Kendler study. Food and water were again available in one or another of the end boxes. Satiated for these objects, the *S*s were motivated to run in the maze by anxiety. Prior to the learning in the maze, the *S*s were trained to run down the stem of the maze, the floor of which was an electrified grid, to escape the shock in an empty end box. Following this preliminary training, 10 days of training (four trials per day) in the maze situation were given in which the *S*s were required to escape by either arm. Shock was not given during this series, the animals presumably being motivated to escape from the maze by fear or anxiety. The experimental question, as in all the above experiments, was whether the *S*s would be able to acquire the appropriate cognitions as to the locus of food and water, while motivated by anxiety and rewarded by escape from anxiety.

Despite an even more thorough attempt to satiate the *S*s for the food and water than had been necessary to prevent eating or drinking in our other investigations, 14 of the 24 *S*s in this experiment either ate or drank or did both. Considering the results of the remaining 10 *S*s that never drank or ate, six responded appropriately on the first test and four on the second, a purely chance result.

In marked contrast to this test performance was that of the *S*s that ate or drank at one time or another in the experiment. Thirteen of these fourteen *S*s responded appropriately on the first test and eleven on the second. The results for the first test are significant beyond the one percent level of confidence. The departure from chance expectation (50 percent correct) for the two tests combined is also significant beyond the one per cent level of confidence. The success of these latter *S*s is rather remarkable in the light of the fact that only about eight reinforcements involving drinking and eating were obtained on the average per *S* in the forty training trials.

Such results as the above, combined with the findings of our earlier study (22) and its confirmation by Walker (28), Kendler (10), Grice (1), and Kendler and Mencher (11), which have shown that rats are unable to acquire either cognitions as to the locus of food when not hungry but under a thirst drive or cognitions with regard to the location of water when not thirsty but motivated by hunger, have led Tolman to offer a modification of his views concerning the acquisition of sign-gestalt-expectations, or as he now prefers to designate them, field expectancies (27). The extent of Tolman's reversal of viewpoint is shown by the following quotation:

> I used to be so impressed by the latent learning experiments of the type invented by Blodgett, in which no reward was introduced during

the learning periods, that I was apt to formulate the conditions involved in such field-expectancy learning primarily in terms of frequency alone and as if motivation played no role. However, if I did this, I was in error. It is obvious that completely unmotivated animals will not learn. They will go to sleep or otherwise divorce themselves from the task. So it must be emphasized that in the Blodgett experiments, even though the animals were not rewarded, they *were* motivated. . . . Summing it up then, it appears that motivation conditions *are* very important for the building up of field expectancies (**27**, p. 150).

While clear and emphatic about the necessity of motivation for learning field-expectations, Tolman is less decisive about the role of reinforcement. He denies the view: "that such learning consists in the stamping in of S-R habits by reinforcements" (**27**, p. 151). But he goes on to state that: "The presence of reinforcement in a particular locus makes that locus a goal which determines what performance will take place but it does not stamp in S-R connections, though it probably does give a special vividness to that locus in the total field expectancy" (**27**, p. 151).

So far as the experimental studies discussed above are concerned, they point just as strongly to the necessity for reinforcement as they do for motivation. In every single investigation some form of reinforcement was present, and in one instance the experiment could not be conducted when no reinforcement was provided. However, none of these studies is able to answer the question as to whether learning can occur in the absence of reinforcement. As described earlier, they are concerned only with the question as to whether the learning of the locus of certain kinds of goal objects can occur under motivational-reward conditions not involving these objects.[3]

[3] There is no implication that these studies necessarily demand a reinforcement interpretation and are opposed to a non-reinforcement interpretation. While they are in line with a reinforcement assumption, they are not critical with respect to the issue, as none of the studies eliminated completely some form of reinforcement.

In terms of Tolman's theory, these experiments attempt to answer the question as to whether the learning involves the formation of cognitions or field expectancies involving the stimulus events at the beginning of each alley and the stimulus events (food or water) at the end of each alley. The perception of these events in sequence is assumed to result in the knowledge (cognition) of the relationship between them. Given these cognitions, plus the knowledges that these objects satisfy their respective primary drives, the choice behavior of the subject will depend upon the relative demand value (Δ) of the alternative significates. Thus, if the S is hungry on the test and satiated for water, and assuming food is at the end of the left alley, then the strength of the tendency or force to the left response (B_L) will be greater than that to the right (B_R). Employing the symbols used by Tolman in his article on secondary demands (**26**), this derivation may be represented as follows:

$$B_L > B_R \text{ if } \Delta_L > \Delta_R$$
$$\Delta_L > \Delta_R \text{ if } \Delta_F > \Delta_W$$

and if hypotheses $H_{L\to F}$ and $H_{R\to W}$ exist,

$$\Delta_F > \Delta_W \text{ if } D_h > D_t$$

and if hypotheses $H_{F\to C_h}$ $H_{,W\to C_t}$ exist.

Where Δ represents secondary demands, such as demand for a particular alley, for food, etc., H represents hypothesis, cognition, equivalence belief, etc., D_h primary demand for consummation of food, i.e., hunger; D_t primary demand for water, i.e.,

thirst, and C_A and C_t consummation of hunger and thirst respectively.

Taking the results of all of the above experiments, the findings are not too favorable to Tolman's S-S theory. In those instances in which the *S*s were motivated for one of the significates, the test data have been uniformly negative. In the experiments in which some motivation other than that for the significates was employed, the findings are ambiguous. In some (Spence, Bergmann, and Lippitt; Meehl and MacCorquodale, and Kendler) one of the tests has agreed with Tolman's theory, whereas a second test has not supported it. In other studies (15, 24) the test data have been uniformly opposed to the implications of Tolman's theory.

Attempting to fall back on the claim that the objective conditions of presentation of the significates in these experiments were not adequate for reception, such as Leeper (14) has done, is refuted by the positive instances, as the latter were precisely the same as those in which the results were negative (33). Furthermore, such conditions as Kendler and Mencher (11) arranged, preclude the possibility that their subjects did not 'see' the food. It is at this point, of course, that Tolman introduces the factor of motivation. He writes, "Thirsty animals apparently do not notice food, even though the experiment be rigged as it was by Kendler and Mencher to seem to force them to notice that the cups which did not contain water did contain food" (27, p. 150). That is, under the motivation of thirst, the animals did not, he claims, notice (perceive?) the food.

But why, we may ask, did the *S*s of our present experiment (accepting the positive results of the first test) apparently notice the food and water while motivated to get to the social cage? And why, when the social goal box was brought nearer to the point of entrance to the alleys, as in Maltzman's study (15), did the rats apparently fail to notice the food and water? Again, what lies behind the failure of the *S*s to notice the food or water in the experiment in which the *S*s were motivated by anxiety and need to escape the anxiety arousing situation? Under just what motivational conditions will animals perceive or not perceive the presence of the significates? It is obvious that Tolman's theory is as yet totally inadequte to account for these experimental results. Much experimental work remains to be done before any satisfactory theory concerning the role of motivation in perception is achieved. Tolman himself is well aware of the primitive, undeveloped nature of his theorizing, even though some of his more ardent followers are not.

We turn now to the S-R interpretation of these phenomena. So far as we know, neither Hull nor any of his followers have attempted to formulate a *complete* theoretical account of even so simple an instance of trial-and-error learning as is involved in the single T-maze. Many of the critics (3, 13, 18) of Hull have either never understood or have forgotten that in his *Principles of behavior* (6) Hull deals only with the basic principles revealed in classical and instrumental conditioning. Hull's basic principles do *not* include all of the principles that are important for all instances of learning, animal and human. By basic he means certain principles revealed in simple instances of learning from which he believes other, just as important, learning principles may be derived. Not treated in his *Principles of behavior* are such highly important principles as that of the habit-family hierarchy (5), the principle of the fractional anticipatory goal response (4) and a number of others.

The fractional anticipatory goal response is particularly important for all

instances of complex learning including the type of problem with which we are here concerned.[4] In the original report of the present experiment (21) and in an unpublished symposium paper presented at the 1941 meeting of the Midwestern Psychological Association (19) an attempt was made to show how this mechanism would operate in the present type of experiment. Before presenting this, however, the necessity for including in the theory some other mechanism than the usual excitatory potentials between the environmental and internal drive cues and the responses may be shown by considering an experiment of Kendler (8). Using a single T-maze, one arm of which had food and the other water, Kendler had his subjects *both* hungry and thirsty during the training series. In the test series only one of the drives, hunger *or* thirst, was operative. Under this test condition Kendler found that his *S*s responded correctly a highly significant number of times. But it is obvious that under the training condition each of the responses was reinforced equally often and hence presumably was of approximately equal strength at the completion of training. How then can one predict the differential strengths under the two test conditions? The explanation suggested by Spence (8, 19) was that the stimulus cues in the water arm and end box become, during training, conditioned to fractional anticipatory drinking acts (r_w) and, in turn, the proprioceptive cues (s_w) resulting from these anticipatory acts become conditioned to the response of entering and continuing locomotion in this alley. In a like manner, conditioned, fractional anticipatory eating responses (r_f) develop in the food alley and these cues (s_f) become conditioned to the behavior of entering this alley. During the test series when only one drive is operative, the anticipatory act related to the goal for which the subject is motivated will, because of the greater strength of the particular drive stimulus, Hull's principle of stimulus dynamism (7), be much stronger (more vigorous) than the other and hence will produce stronger proprioceptive cues. Thus, if the *S* is thirsty, proprioceptive cues from anticipatory drinking responses will be stronger than those from anticipatory eating. As these cues (s_w) will be conditioned to the response of entering the alley leading to water, they will tend to give this response the greater excitatory strength (stimulus dynamism). When the animal is hungry, the proprioceptive cues (s_f) would be stronger and this would give the response of entering the food alley the greater excitatory strength. In other words, the relative excitatory strengths of the competing responses will be a function of the relative strengths of the two fractional anticipatory goal responses, and the latter will be dependent upon which drive is present and which is absent.

The application of this principle to the experiment of the present study is similar. In our experiment, however, the fractional anticipatory goal responses are much weaker as the *S*s do not actually eat or drink during the training period. The sight of the food (or water) evokes conditioned fractional eating (drinking) responses. These, as described above, become conditioned to the cues of their respective alleys, the reinforcement being provided by the social goal, or possibly, as in higher order conditioning, by the sight of the food (or water) itself (i.e., secondary reinforcement). The much greater weakness of these fractional goal responses in this experiment as compared with the Kendler experiment, in which eating and drinking actively occurred should lead to much poorer conditioning, and hence, presumably smaller differences between the excitatory tendencies of their proprioceptive cues. This would nicely account for the poor test results obtained in the present experiment as compared with Kendler's motivated *S*s.

Making use of the same explanatory schema, Maltzman (15) has suggested that the explanation of the still poorer results he obtained was possibly due to

[4] As one of us (20) has pointed out, the fractional anticipatory goal response is an instance of classical conditioning which itself enters into and plays a role in all the more complex types of learning, including instrumental conditioning.

greater interference with the fractional anticipatory eating and drinking responses by the stronger, anticipatory social goal responses. According to this view, the stronger the motivation operative in the training situation and the stronger the correlated goal response, the poorer will be the test results, as these responses will interfere with the fractional eating and drinking responses. Another implication of this theoretical picture is that the more highly differentiated the fractional anticipatory goal responses elicited by the significates, the better will be the results on the tests.

The failure of the *S*s on the second test in the present investigation and in the similar study of Kendler (10) would be accounted for by the fact that all of the *S*s of our study and those correct on the first test in the Kendler experiment were rewarded on the first test day. This relatively strong reinforcement under motivation for the significate or goal object would tend to strengthen the tendency to enter this alley, thus upsetting the balance of strength in favor of response to this side. As the second response was always to the opposite side there was thus introduced a factor that would be unfavorable to success on this test.

This explanation does not hold, however, in the case of Meehl and MacCorquodale's study, as the goal object for which the *S* was motivated on the first day's test was not present. Hence, the response to this side would not receive this relatively strong reinforcement, but only the usual reinforcement provided by escape from the maze to the social cage. It may or may not be significant that the shift from the training performance to the appropriate test side was greatest in this study even though not significant.

Finally, the completely negative test findings in the studies of Grice (1), Walker (28) and Kendler and Mencher (11), in which the *S*s were thirsty during training and found water at the end of each pathway and food in one can be easily explained in terms of the concept of the fractional anticipatory goal response. In these experiments the drinking response would be dominant in both pathways and hence would become conditioned to the cues at each alley entrance. This explanation assumes, of course, that the drinking and eating responses are incompatible—a not unreasonable hypothesis.

It is readily apparent that this S-R theory, like the rival sign-gestalt theory of Tolman, is much in need of more specific, detailed elaboration. Even in its present stage of development, however, it is more detailed and definitive than that offered by Tolman. It has, at least, suggested some of the variables, e.g., the various motivations, their relative strengths, the relations among the several goal responses, etc., that are important determiners of the behavior in the situation. Tolman's original formulation recognized none of these variables and even in his most recent formulation he has only vaguely referred to the role of motivation. One may suspect that when the sign-gestalt theorists get around to making some definite assumptions, they will become more and more similar to those suggested above.

Summary

The present study is one of a series of investigations undertaken in the Iowa laboratory to test Tolman's sign-gestalt theory of simple trial-and-error learning. The *S*s, 39 white rats, were given seven days of experience (four trials per day) in a single T-maze in which one path led to water and one to food. The motivational-reward condition employed to get the animals to run in the maze during the training period was that of getting to a social cage containing a cage mate. In the first of two test series given following the completion of training, 23 of the *S*s were run under thirst while satiated for food and 16 were hungry and satiated for water. In a second test on the following day the motivational conditions for each *S* were reversed.

The results of the first test showed a significant shift from the pre-test choice of the *Ss* to the alley containing the goal object for which they were motivated. The results of the second test, on the other hand, were entirely negative, with little or no shift in the direction of the appropriate alley.

The experimental findings are discussed in relation to both the cognitive theory of Tolman and an S-R theory that makes use of the fractional anticipatory goal response.

(Manuscript received October 20, 1949)

References

1. Grice, G. R. An experimental test of the expectation theory of learning. *J. comp. Psychol.*, 1948, **41**, 137–143.
2. Hayes, K. J. An experimental criticism of the Spence and Lippitt procedure. *Amer. Psychologist*, 1949, **4**, 223.
3. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century Crofts, 1948.
4. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38**, 487–506.
5. Hull, C. L. The concept of the habit family hierarchy and maze learning. *Psychol. Rev.*, 1934, **41**, 33–52, 134–152.
6. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
7. Hull, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–77.
8. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
9. Kendler, H. H. An investigation of latent learning in a T-maze. *J. comp. Psychol.*, 1937, **40**, 265–270.
10. Kendler, H. H. A comparison of learning under motivated and satiated conditions in the white rat. *J. exp. Psychol.*, 1947, **37**, 545–549.
11. Kendler, H. H., & Mencher, H. C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, **38**, 82–88.
12. Leeper, R. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
13. Leeper, R. Dr. Hull's *Principles of behavior*. *J. genet. Psychol.*, 1944, **65**, 3–52.
14. Leeper, R. The experiments by Spence and Lippitt and by Kendler on the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, **38**, 102–106.
15. Maltzman, I. A study of learning under an irrelevant need. Unpublished M.A. thesis, Univ. Iowa, February, 1948. Also *Amer. Psychologist*, 1949, **4**, 352.
16. McNemar, Q. *Psychological statistics.* New York: John Wiley & Sons, Inc., 1949.
17. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. Psychol.*, 1948, **41**, 372–396.
18. Ritchie, B. F. Hull's treatment of learning. *Psychol. Bull.*, 1944, **41**, 640–662.
19. Spence, K. W. Symposium: learning as related to need and the subsequent motivation of such learned behavior. *Psychol. Bull.*, 1941, **38**, 721.
20. Spence, K. W. Theoretical interpretations of learning. Chapter 18 in: *Handbook of experimental psychology.* New York, John Wiley & Sons, Inc. (In press)
21. Spence, K. W., & Lippitt, R. 'Latent' learning of a simple maze problem with relevant needs satiated. *Psychol. Bull.*, 1940, **37**, 429.
22. Spence, K. W., & Lippitt, R. An experimental test of the sign-gestalt theory of trial and error learning. *J. exp. Psychol.*, 1946, **36**, 491–502.
23. Spence, K. W., & Kendler, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, **38**, 106–109.
24. Thune, L., Dusek, R., & Spence, K. W. Learning under relevant and irrelevant needs. *Amer. Psychologist*, 1949, **4**, 352.
25. Tolman, E. C. Theoretical interpretation of learning. Chap. 12 in: Moss, F. A., *Comparative psychology*, New York: Prentice-Hall, 1934.
26. Tolman, E. C. Demands and conflicts. *Psychol. Rev.*, 1937, **44**, 158–169.
27. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, **56**, 144–156.
28. Walker, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, **38**, 39–49.
29. White, R. K. The case for the Tolman-Lewin interpretation of learning. *Psychol. Rev.*, 1943, **50**, 157–186.

AN EXPERIMENTAL INVESTIGATION OF TRANSPOSITION AS A FUNCTION OF THE DIFFERENCE BETWEEN TRAINING AND TEST STIMULI[1]

BY TRACY SEEDMAN KENDLER

Psychological Laboratory, State University of Iowa

In the usual discrimination problem, two or more members of a stimulus dimension, e.g., size or brightness, are presented to the subject. Response to one of them is consistently rewarded while response to the other either receives no reinforcement, or is punished. After a number of trials the subject learns to choose one of the stimuli consistently. The question of whether this learning is based on a tendency to respond to specific stimulus aspects or to the relational features of the stimulus complex, e.g., brighter or larger, has been of considerable theoretical importance.

The Gestalt psychologists have naturally been interested in demonstrating that the relational aspects of learning are the most significant. They have emphasized the fact that experimental studies have consistently shown that after subjects have learned the discrimination problem, e.g., have learned to choose the larger of two stimuli and are presented with the originally positive stimulus and a still larger one, they now respond more often to the new and larger stimulus than to the one that was originally positive. Such apparent response to the relational aspect of the stimulus complex has been interpreted by the Gestalt psychologists as irreconcilable with associationistic and stimulus-response theories.

In 1937, however, Spence (7) proposed a theoretical scheme based on stimulus-response principles to explain such 'relational' responses. Utilizing the concept of generalization as revealed by conditioning experiments, plus a theoretical formulation of discrimination learning (6), he was able to explain such behavior without assuming that the response was to the relational aspects of the situation. His theory is based on the following five assumptions:

1. As a result of successive reinforcements, the positive stimulus of a discrimination combination acquires an excitatory tendency of a given amount to the response of approaching it, etc.

2. There is a generalization of this tendency along the stimulus continuum and it takes the following form: $E = Ma^{-cX^d}$ (where X is the distance in logarithmic units between training and test stimuli).

3. Failure of reinforcement results in the development of an inhibitory tendency of a given amount for the non-reinforced stimulus.

4. This inhibition generalizes along the stimulus dimension according to the same general formula as does the excitation.

5. The effective excitatory strength of any stimulus member of the dimension is the algebraic summation of these excitatory and inhibitory tendencies.

The implications of this theory can be most readily seen from an examination of Fig. 1, a diagrammatic representation taken from one of Spence's later articles (8). The subject is assumed to have been trained to choose a stimulus square of 160 sq. cm. in preference to one of 100

[1] The present paper constitutes the major portion of a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in the Department of Psychology at the State University of Iowa in 1943. The writer wishes to thank Professor Kenneth W. Spence for his invaluable aid in the direction of the investigation. Appreciation is also extended to Dr. David Ehrenfreund for his untiring cooperation in the construction of the apparatus.

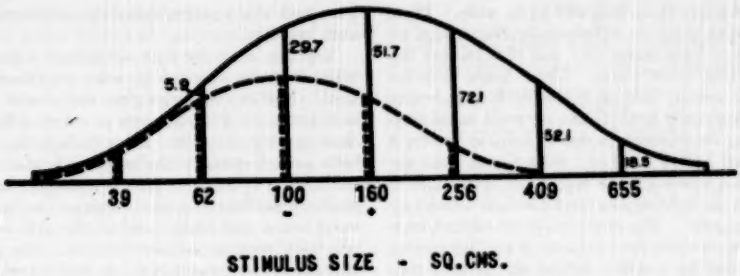


FIG. 1. Diagrammatic representation of the relations between the hypothetical generalization curves, positive and negative, after training on the stimulus combination 160(+) and 100(−)

sq. cm. According to the theory, if stimulus 160 is presented with stimulus 256, the effective excitatory strength of the new stimulus is greater than the original positive one and therefore the animal would be expected to respond to the new and larger stimulus.

Although most infra-human organisms respond in a manner more or less consistent with the direction of the original training, the percentage of such responses to the transposed pair of stimuli is often fairly low and sometimes even in an opposite direction. It is important for any theory of transposition to be able to account for such variations. Spence's theory would suggest that one of the significant variables determining the proportion of choices consistent with original training in such transposition tests is the distance of the test pair from the original training pair. Experiments by Gulliksen (2), Klüver (5), Spence (7) and Flory (1) have obtained evidence which give definite support to this general implication. Thus when the amount of transposition (percentage of responses consonant with original training) on test pairs near the original training pair is compared with the amount occurring on test pairs farther removed, the latter has almost invariably been smaller.

None of the previous experimental studies, however, has provided sufficient data to permit a description of the relationship holding between the amount of transposition and distance along the stimulus dimension between the training and test pairs. The present investigation was aimed specifically at a determination of the precise nature of this law in the case of brightness discrimination with rats, and also a study of the implication of the results for a theory of transposition.

EXPERIMENTAL PROCEDURE

Apparatus

The apparatus used was a modification of one employed previously by Flory (1) and Spence (9). Its principal parts are shown in the diagram in Fig. 2. A turntable, divided into two platforms, A1 and A2, by means of partitions, stood about four feet above the floor and moved on a pivot made by two iron pipes. The smaller pipe, which was fastened to the table, fitted closely inside a larger one which rose from a base on the floor. A one-foot alley (B), two in. in width, connected the two open sides of the table and a vertically sliding door in the middle of this alley permitted *E* to control the subject's passage from one platform of the apparatus to the other.

The turntable was utilized in the following manner: At the beginning of each day's trials, *S* was placed on the far platform of the apparatus (A2 in diagram) and then the door was raised, permitting access to the platform adjoining the stimulus pathways. After making its choice the animal was encouraged to go back on to the turntable. *E* then rotated the apparatus so that the subject was again on the platform distant from the stimulus objects. Upon the appropriate adjustment of the stimulus objects, the next trial ensued.

The second main portion of the apparatus, namely the stimulus objects (brightnesses) to be discriminated consisted of two glass pathways

(C1 and C2) 9½ in. long and 4½ in. wide. These pathways could be differentially illuminated by means of light boxes (H1 and H2) located immediately below them. These boxes were arranged side by side on a two by four and were placed directly beneath the plywood board supporting the stimulus paths. Identical apertures (J1 and J2) 7½ × 4½ in., were cut in both the plywood board and the top of the light boxes to permit the light to pass from the light sources up to the paths. The boxes could be rotated from one side to the other by means of a pulley system controlled by a wheel behind the screen, thus permitting the position of the stimuli to be changed from trial to trial. The light source for each path was an electric bulb at the bottom of each light box.

Eight in. from the floor of the box, a plate of milk glass was inserted in order to diffuse the light. One in. above this glass was a small shelf designed to support shutters in which different sized openings were cut. After the light from the bulb passed through the lower milk glass and the circular opening, it filtered through a cut of ground glass that covered the aperture in the plywood board and finally reached the path which was itself constructed of milk glass. The glass and the empty spaces through which the light filtered served to diffuse the light evenly over the pathway and keep it from getting over-

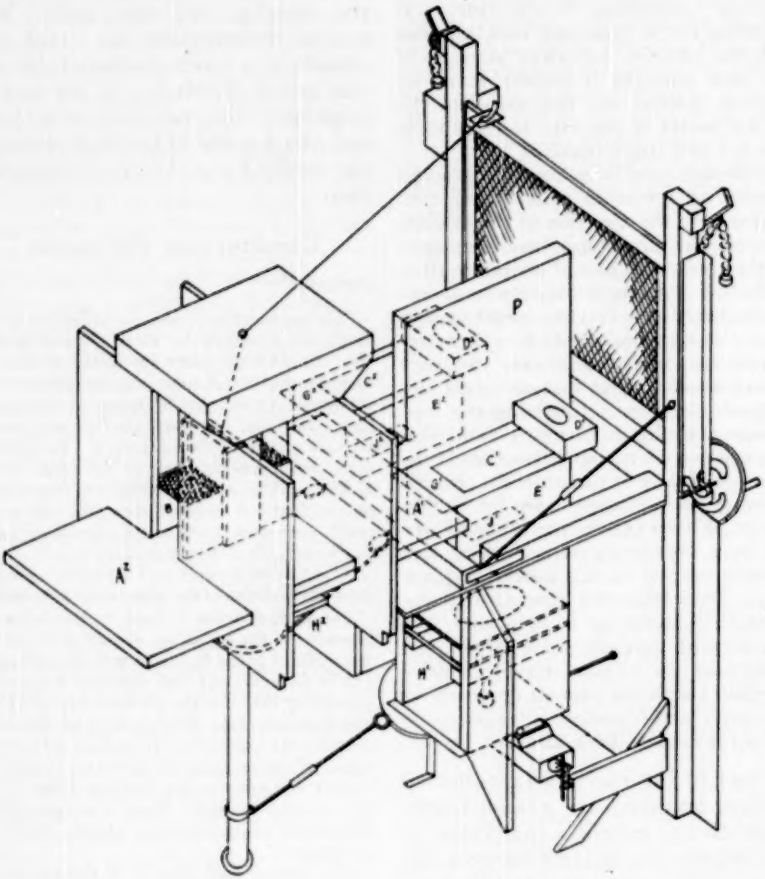


FIG. 2. Isometric diagram showing the essential features of the apparatus

heated. All portions of the apparatus that were visible to the rat, except the stimulus paths, were painted a flat black.

The brightness values of the stimulus paths were controlled and manipulated by: (1) the size of the circular openings, (2) the wattage of the bulbs, and (3) the voltage impressed upon them. Using these devices in the various combinations indicated in Table I, eight brightness values ranging from .011 to 204.7 ft. c. were obtained. Brightness is here defined as the measure in apparent foot candles obtained with the use of the Macbeth Illuminometer. An attempt was made to have each successive stimulus value approximately five times brighter than its predecessor, thus providing for stimulus pairs with a constant ratio between the two brightnesses. As may be seen from Table II however, this ratio was not quite constant, but varied from 4.4 to 5.9.[2]

Stimulus conditions in learning and test series

Two different conditions of original training were used. Under one (Condition I) the subjects (N = 58) were trained to choose the brighter member of stimulus pair No. 0 in Table II, i.e., stimulus .063 in preference to stimulus .011. Following attainment of the criterion of learning each subject was given a series of 10 test trials

TABLE I

COMBINATION OF WATTAGE, VOLTAGE AND AREA OF CIRCULAR OPENINGS USED FOR EACH BRIGHTNESS STIMULUS

Brightness in Apparent Foot Candles	Wattage	Voltage	Area of Circular Opening in Sq. Cm.
.011	15	100	3.14
.063	40	100	12.6
.370	60	100	22.9
1.62	100	100	55.4
8.49	100	100	358.7
40.46	150	100	615.8
41.22*	150	130	318.5
204.74*	300	130	No shutter

* For these two stimulus brightnesses both pieces of filtering glass were removed.

[2] According to the photometric measures obtained just prior to running the first subjects, the ratios were actually five to one. When the measures were repeated toward the close of the experiment, the values deviated somewhat from the original measures. It is the second set of measures that appears in Tables I, II, and III.

TABLE II

STIMULUS PAIRS USED IN CONDITION I

Stimulus Pair	Brightness in Apparent Foot Candles		Ratio
	Positive	Negative	
0	.063	.011	5.7:1
1	.370	.063	5.9:1
2	1.62	.370	4.4:1
3	8.49	1.62	5.2:1
4	40.46	8.49	4.8:1
5	204.74	41.22	5.0:1

with *one* of the six stimulus combinations shown in Table II. The subjects were distributed among the various test groups in a random manner; all groups contained 10 subjects except the group which was tested on the stimulus combination employed in the original learning. There were eight subjects in this latter group.

In a second condition (Condition II) the subjects were trained positively to the darker pathway of stimulus pair No. 0 in Table III.[3a] After learning this discrimination, these subjects were tested with *one* of the darker stimulus combinations. To each of the four test groups six subjects were assigned at random.

In the first condition, then, training was to the brighter of a pair of stimuli well down in the brightness scale and the extent of transposition to pairs of brighter stimuli of approximately the same ratio was measured. In the second condition the subjects were trained initially to the darker of two relatively bright stimuli and transposition tests were given on still darker stimulus pairs. We can speak of the transposition tests being in a positive direction, by which is meant that they are in the direction of the positive member of the original stimulus pair. We can also measure the extent of the distance from the original training pair in terms of the successive step intervals, the ratios of which are approximately five to one.

Training and test procedures

Following a period of several days in which the subjects were fed their daily diet of 5½ gm. of Purina Dog Chow at a regular hour and were accustomed to handling, preliminary training on the apparatus was begun. On the first day the subject was placed on the apparatus and per-

[3a] The brightest stimulus combination could not be employed as a training pair because the apparatus overheated when utilized in a long learning series.

TABLE III

STIMULUS PAIRS USED IN CONDITION II

Stimulus Pair	Brightness in Apparent Foot Candles		Ratio
	Positive	Negative	
0	8.49	40.46	4.8:1
1	1.62	8.49	5.2:1
2	.370	1.62	4.4:1
3	.063	.370	5.9:1
4	.011	.063	5.7:1

mitted to explore it. On the two following days, 10 small food pellets were placed in each of the food troughs at the end of both stimulus paths (D1 and D2) and the subjects were permitted to find and eat them. During these trials both pathways had the same brightness value, that of the positive stimulus of the original training pairs. Additional general illumination was provided on the first of these two days by a light in the box shown at the top of the turntable. The only illumination on the last day of preliminary training and all subsequent training and test days was that provided by the stimulus paths. During the training and test series, the rats were kept in the dark room for 15 min. before they were placed on the apparatus to permit dark adaptation to occur.

The learning series, consisting of 10 trials per day, was continued until the criterion of learning was attained. The criterion was 18 out of 20 correct trials in both Conditions I and II, except that the further requirement that the last 10 trials be correct was adhered to in Condition I. Because the discrimination problem proved to be much more difficult in Condition II, the added requirement that all of the last 10 trials be correct was dropped. A correct response was rewarded by a small pellet of food (.2 gm.) while failure of such reinforcement constituted the sole consequence of responding incorrectly. The position of the positive pathway was varied according to the following schedule: RLLRRL-RLLR or the reverse of this order.

In the test series each animal was tested on one stimulus combination only. Ten trials were given and responses to either stimulus pathway, darker or lighter, were rewarded regardless of the original learning conditions.

Subjects

In all, 82 hooded rats were used as subjects. Fifty-eight animals, half of which were males, were trained under Condition I; 24 subjects, 13 males and 11 females, were trained under Condition II. All animals were from the colony maintained by the Department of Psychology, State University of Iowa. They ranged in age from 75 to 145 days at the beginning of training.

RESULTS

Acquisition of discrimination habit

The initial discrimination problem proved to be a difficult one to learn, especially in the case of the subjects trained under Condition II, i.e., to the darker of the stimulus pair. The latter required an average of 446 trials to reach their more lenient criterion while the subjects that learned under Condition I took, on the average, 252 trials to learn. This difference in difficulty of the two learning situations suggests the possibility of a preference for the brighter stimulus.[3] Another factor that possibly contributed to the difference is the fact that the ratio of the brightnesses was somewhat smaller in Condition II (4.8:1) than in Condition I (5.7:1).

Transposition test results

It has generally been accepted that the best measure of transfer effects is provided by the first test trial, since all subsequent trials in the test series are likely to reflect the effects of the test trials. It is, however, infrequently applied because of the large number of subjects required for reliability. Most studies apply a less stringent measure, such as the percent of 'correct' responses in a series of five, ten, or more test runs. In the present research three measures were planned and a fourth one emerged from the results. The three planned measures consisted of the percent of responses

[3] In an exploratory study preceding the present investigation, a slight but statistically significant (five percent level) preference for the brighter member of a stimulus pair was found. The subjects, 20 hooded rats, were each given 20 trials which were rewarded regardless of which stimulus they chose.

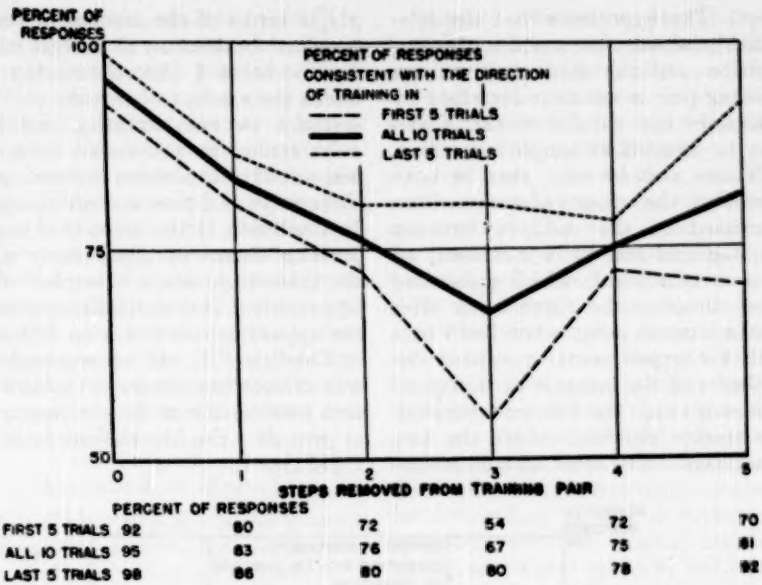


FIG. 3. The percent of responses consistent with the direction of training for subjects trained to the brighter stimulus and transposed up the brightness scale (Condition I)

consistent with the direction of training in: (1) the initial test trial, (2) the first five test trials, (3) 10 test trials. The fourth measure was composed of the percent of transposed responses in the second five test trials.[4]

The results on the initial test trial are presented in Table IV. Because of the small N the results appear quite unstable.

The percent of transposed responses for the first five test trials, the second five test trials and all 10 trials are presented in Figs. 3 and 4.

[4] In Condition I the subjects were trained to the brighter stimulus. A response 'consistent with the direction of training,' or a 'transposed response,' during the test trials would be the choice of the brighter stimulus of the test pair presented. For Condition II subjects, who were trained to the darker stimulus, a response 'consistent with the direction of training' would be towards the darker stimulus of the test pair presented.

An examination of the transposition results presented in Figs. 3 and 4 reveals that, regardless of which measure is invoked or which experimental condition is described, there was a tendency for the subjects to respond in the direction consistent with the direction of training at every step tested. The percent of 'consistent' responses was, however, less at each of the transposed steps than it was at the zero (training)

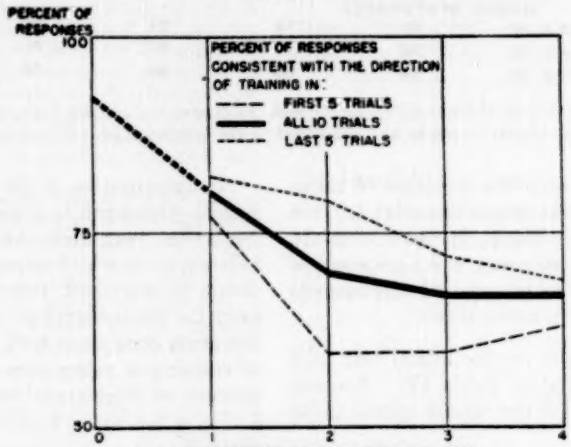
TABLE IV

PERCENT OF SUBJECTS WHOSE INITIAL TEST RESPONSES WERE CONSISTENT WITH THE DIRECTION OF TRAINING

	Steps Removed from Training Pair					
	0	1	2	3	4	5
Condition I	88	80	70	60	80	67
Condition II		67	33	67	50	

Step. The hypothesis that the relationship between the amount of transposition and the distance from the training pair is suitably described by a straight line parallel to the X-axis may be rejected by simple inspection.

It can also be seen that in both conditions the amount of transposition decreased as the distance between training and test pairs increased, up to a certain point; which point and what direction the curve takes after that minimum point is reached varies with the experimental condition described and the measure of transposition involved. The difference between the results obtained under the two conditions can be dealt with most simply in terms of the amount of transposition obtained on all 10 test trials. In Condition I the relationship between the amount of transfer and the distance between training and test pairs resembles a U-curve, with the amount of transposition decreasing for three steps and then increasing again. In Condition II the amount of transposition decreased consistently until the third step, where it leveled off at 67 percent. Due to the limitations of the apparatus, there was no fifth step in Condition II, and consequently no way of knowing whether an additional step would continue the previous trend or provide a rise like the one found in Condition I.



PERCENT OF RESPONSES

FIRST 5 TRIALS		77	60	60	63
ALL 10 TRIALS	92	80	70	67	67
LAST 5 TRIALS		82	79	72	69

FIG. 4. The percent of responses consistent with the direction of training for subjects trained to the darker stimulus and transposed down the brightness scale (Condition II).

(The heavy dotted line represents an estimate of the subjects' performance on Step 0 during the 10 test trials. The value assigned to this step is the percent of correct runs made by the 24 subjects in the 20 criterion runs. This procedure was utilized because the following results were obtained in Condition I: Eight animals, in this latter condition, that were trained on Step 0 were actually presented with the same pair of stimuli for 10 test trials. The percent of correct responses for these subjects on the 20 criterion trials was 94; the percent of correct responses on the 10 test trials was 95. Since there was such similarity between the results obtained in the criterion and test trials on Step 0 in Condition I, it appeared justifiable to estimate the corresponding value in Condition II as described above.)

Another salient result to be noted is the increase in the amount of transposition as the test trials proceed. That is, when the percent of transposition obtained in the first five test trials is compared with the percent of transposition in the second five trials, the latter is greater at every experimental step in both conditions. A difference in the same direction that is repeated on 10 independent points would occur by chance alone less than one time in a thousand. It therefore appears reasonable to accept the difference in the amount of transposition between the two measures (first and second five trials) as reliable.

Discussion of Results

One of the purposes of the present investigation was to ascertain the nature of the law relating the percent of responses in the transposition tests consistent with the original training, to the difference between the training and test stimulus pairs. Ordinarily the next step in an investigation of this sort is to fit a mathematical function to the empirical results which relates the dependent to the independent variable. The experimental results have shown, however, that the nature of this relationship differs somewhat (especially in Conditon I) depending on which measure of transposition is used. The first trial appears, in this research, to be too unstable a measure on which to base a functional relationship. Analysis of the results obtained on all 10 test trials reveals that the last five test trials contain significantly more transposition than the first five trials. It is clear that in ascertaining the empirical function sought it becomes desirable to determine which of the various measures most directly reflects the variables that the investigation was primarily concerned with.

It would ordinarily appear that the first five trials would be least influenced by extraneous factors; and that the rise in transposition in the later stages of the test series was due to the reinforcements received on the previous test runs. Fortunately, this hypothesis could be tested in the following manner: The data were scanned for subjects that responded more frequently during the first five test trials in the direction that was *not* consistent with the direction of training. If the reinforcements that occurred during the test trials were a significant factor it would be expected that these animals would show a decrease in transposition in the second five trials since they had received more reinforcements for responding in the direction *not* consistent with the original training. There were three subjects, both conditions combined, that transposed only once in the first five trials. In the second five trials their mean transposition was 3.0. Similarly, there were 13 subjects that transposed only twice in the first five trials. Their mean transposition in the second five trials was 3.6. Apparently there was a rise in transposition for these 16 rats in spite of receiving more reinforcements for responding in a direction *not* consistent with their original training.

If the greater percent of transposition in the second five trials is not attributable to learning during the test trials, a more satisfactory explanation of the discrepancy might lie in the fact that the amount of transposition in the first five trials was somehow depressed. There are several possible explanations for how this might have occurred.

One explanation involves consideration of the light adaptation of the subjects. It is relevant to repeat here that the experiment was performed in a dark room, the stimulus paths providing the only

illumination. The subjects were kept in the dark room for 15 min. before they were placed on the apparatus each day to permit dark adaptation to occur. There were no data on dark adaptation in rats, and therefore this time was a rough estimate. The consistent increase in transposition, found even at the zero (training) step, suggests the possibility that a certain amount of consequent exposure to the illumination level of the brightnesses to be discriminated was required for the adaptation mechanisms of the eye, light or dark, as the case may be, to reach their maximum efficiency. If this be true, the initial trials would produce less transposition than the later trials as a consequence of poorer discriminability.

Another possible explanation would be that the test stimuli, since they had not been previously experienced, acted as external inhibitors for all responses, tending more to chance reaction. As the test trials continued, the new test stimuli gradually lost their inhibiting power and therefore the amount of transposition increased.

Since this study was not designed for this purpose, the data do not lend themselves to a test of the validity of these alternative explanations (except for the elimination of the reinforcement hypothesis). Therefore, although the explanations proferred imply that the latter half of the test series provides a less contaminated measure, there is no way to prove it. Moreover, since there are no reports in the literature of transposition of a similar phenomenon, it remains to be seen whether this result is an artifact of the present experimental procedure or a phenomenon common to all test series with new stimuli.

Mathematical functions fitted to the transposition curves are not presented for two reasons: (1) it is not clear which of the measures is least influenced by extraneous factors in the test trials; (2) none of the measures lends itself to a simple, reasonable mathematical description for both experimental conditions.

We turn now to a second purpose of this investigation, the relation of the results obtained to the theory of transposition proposed by Spence (7). The theory postulates that the differential tendency to choose one stimulus over the other is a function of the difference between the generalized effective excitatory strengths of the two stimuli. Attention should be called to the fact that Spence, due to the lack of sufficient data, has not attempted to make specific postulates about the nature of stimulus generalization curves under varying experimental conditions. He did make the general prediction, so far as the present variable is concerned, that the amount of transfer will decrease as stimulus pairs farther removed from the test pair are employed in transposition tests. Actually even this general prediction is not independent of the specific generalization curves involved.

Spence has himself pointed out to the writer, in private communication, the following implications of his theoretical schema (Fig. 1) when examined in detail. In this figure the values appearing within the curves represent the amount of effective excitatory strength accruing to each stimulus. The tendency to choose one stimulus in favor of the other is proportional to the size of the difference in excitatory strengths between them. Beginning with the difference between the original training pair, we see that these values for the successive pairs of stimuli are as follows: 22.0, 20.4, −20.0, −33.6, −18, −14, 0, 0, etc. That is, according to this schema, the percent of responses consistent with the direction of the original training should at first decrease until they drop to well below the 50 percent level (i.e., lead to a reversal of the originally learned preference) and subsequently rise and level off at

a response value, presumably 50 percent, where the difference between the theoretical effective excitatory strengths becomes zero.

The empirical transposition curves obtained from the present experiment do decrease at first, but they do not drop either to or below the chance level. In Condition I the percent of transposition in 10 trials decreases for steps 1, 2, and 3 to 83, 76, and 67 percent, respectively. But at steps 4 and 5 the curve rises to 75 and 81 percent, respectively. These results correspond with the expectations from Spence's schema insofar as there is a decrease in transposition and a subsequent increase. The change in direction predicted from the schema should, however, take place below the chance level. The present results are all above chance. In Condition II, the curve representing the percent of transposition in 10 trials decreases for steps 1, 2, and 3 as follows: 80, 70, and 67 percent. At step 4 the transposition is the same as at step 3, 67 percent. That is, there is no reversal and the curve appears to be levelling off at a value well above 50 percent.

These data do not appear to corroborate all the expectations from the set of theoretical generalization curves utilized by Spence to illustrate his theory. From this we may deduce either that the theory that transposition is a function of the difference between the generalized excitatory strengths is incorrect or, what is more likely, that the type of generalization curves depicted in Fig. 1 are not applicable to the present experimental situation.[6] A better test of Spence's theory will be made when more is known about the nature of stimulus generalization.

[6] There is another possible explanation of the present results. It would involve positing that there are other factors in the present training or test situations that act in addition to those posited by Spence to determine the final results.

It should be pointed out that if Spence's theory should ultimately be proved inadequate, it would not mean that transposition cannot be explained in stimulus-response terms. There is no requirement in modern behaviorism that the stimulus be a purely elemental aspect of the environment. In fact Hull takes pains to overcome "the misleading suggestion of singularity and simplicity otherwise likely to be conveyed by the *S* of the symbol $_{S}H_{R}$" (3, p. 206). He devotes a chapter in this book to a discussion of the conditions under which the patterning of stimulus compounds takes place.

It is possible to describe the present results within the framework of the Hullian theory. The simplest description would be that there was some habit strength built up to respond to the pattern presented by the training pair of brightnesses and that the transposition results obtained represent the generalization along the brightness dimension of this habit strength. It may well be, however, that the situation is considerably more complex, i.e., while the habit strength to the pattern is being built up there are, as Spence postulates, some excitatory and inhibitory strengths being built up to the individual training stimuli. A third factor, the tendency for responses to be greater to more intense stimuli, recently termed stimulus intensity dynamism by Hull (4), is probably active here. If this were the case, the results obtained would represent a combination of the generalization from the above two types of learning plus the stimulus intensity dynamism.

A more precise explanation of the present results may have to wait until there are more data available about the nature of generalization of both individual stimuli and stimulus compounds and, perhaps also, the way that these two types of generalizations interact.

Summary

The purpose of this investigation was to determine the nature of the functional relationship existing between the amount of transposition

and the distance between the pair of test stimuli and the original pair of training stimuli.

Hooded rats ($N=82$) were trained to choose one of two differentially illuminated pathways to food. After the criterion of learning was reached, each subject was presented with one of a series of different pairs of illuminations at varying distances (steps) along the brightness dimension from the training pair. They were given 10 test trials.

In Condition I, the subjects ($N=58$) learned to choose the brighter of two relatively dark pathways and were transposed up the brightness scale. The subjects in Condition II ($N=24$) learned to choose the darker of two relatively bright pathways and were transposed down the brightness scale. The ratio between the brightnesses of the stimuli that composed the test and training pairs in both conditions was approximately 5 to 1.

The results showed that, in Condition I, transposition decreased as a function of the difference between training and test stimuli for three steps. There was a reversal of the tendency for the fourth and fifth step, i.e., the amount of transposition increased. In Condition II the amount of transposition decreased as the distance between the training and test stimuli increased until the third step, where it levelled off. An analysis of the test trials indicated that there was more transposition in the last half of the test series than in the first half.

The empirical results could not be readily fitted to a simple, reasonable mathematical function. The results are discussed in relation to Spence's theory of transposition.

(Manuscript received August 22, 1949)

References

1. Flory, R. M. A study of the factors determining discrimination of size by the white rat. Unpubl. Master's thesis, Univ. Virginia, 1938.
2. Gulliksen, H. Studies of transfer of response, 1. Relative versus absolute factors in the discrimination of size by the white rat. *J. genet. Psychol.*, 1932, **40**, 37–51.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century Co., 1943.
4. Hull, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
5. Klüver, H. *Behavior mechanisms in monkeys.* Chicago: Univ. of Chicago Press, 1933.
6. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
7. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol Rev.*, 1937, **44**, 430–444.
8. Spence, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 257–271.
9. Spence, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.

THE EFFECT UPON GENERALIZED DRIVE STRENGTH OF EMOTIONALITY AS INFERRED FROM THE LEVEL OF CONSUMMATORY RESPONSE [1]

BY ABRAM AMSEL

Newcomb College, Tulane University

AND

IRVING MALTZMAN

University of California at Los Angeles

Introduction

If one postulates, as Hull (4) does, supported by the data of Perin (8) and Williams (11), that excitatory potential and hence response strength is some joint multiplicative function of the learning variable (strength of habit) and the motivational variable (drive strength), then it must follow that any increase in the strength of drive will increase the strength of ongoing responses.

The definition of drive (D) concerns us most here. This is a construct representing the total generalized drive strength. It is thought of as a nonspecific motivational variable, contributed to by all the separate need states which are present in an organism at any moment. Any single need state can be strictly defined by the particular set of antecedent circumstances which give rise to it. Need states, which may contribute drive strength to the total generalized drive condition, are classed as either relevant or irrelevant, a relevant need being one which is reduced by the response being observed in a particular experimental situation, an irrelevant need being one which is not so reduced. The formula suggested by Hull for estimating the total drive strength from the combination of relevant and irrelevant needs is:

$$D = \frac{\dot{D} + \bar{D}}{\dot{D} + M_D},$$

where $\dot{D}$ represents the drive strength contributed by irrelevant needs, $\bar{D}$ the drive strength contributed by the relevant need, and M_D represents the maximum motivational strength. According to the above formula, any additional irrelevant need introduced in an experimental situation will add to the total drive strength unless the relevant need contributes, by itself, a drive strength equal to the maximum (M_D).

On the basis of the foregoing it would be expected that, ordinarily, any incident or event, in an experimentally controlled situation, which introduces an additional need state not necessarily relevant to the response being observed, will increase the strength of that response. This has not, however, been found to be universally true. Several studies have shown that the strength of a previously learned response in a particular situation is diminished by the addition of an emotionally derived need to the existing motivational complex.[2]

[1] The experimental portion of the study here reported was carried out in the Psychological Laboratory of the State University of Iowa.

[2] The distinction will be made here and subsequently (2) between *appetitional* needs, which characteristically result from deprivation of a needed substance, and *emotionally derived* needs which are related to noxious stimulation. Three

Estes and Skinner (3) set up a lever pressing response in a Skinner box with grid floor, using rats as subjects. Hunger was the relevant motivation. Subsequently the animals were shocked in the same situation immediately following the three-min. presentation of a tone. After several tone-shock combinations, it was observed that the presentation of the tone resulted in reduction of the rate of lever depression. In this instance the addition of anxiety to the motivational complex decreased the strength of a previously learned response, though the relevant hunger motivation was still present.

An experiment of Mowrer and Viek (7) is also pertinent to the present discussion. They found that a response of eating in a particular situation was inhibited during a period when the subjects were anticipating shock, and that greater inhibition was associated with what they took to be a more intense fear state. Here again, the presence of fear (anxiety) as an additional need state did not increase the strength of the response being observed, but rather decreased it.

In contrast to the two studies cited and that of Kendler (5), which showed the weakening of a bar-pressing response with the addition of a strong irrelevant thirst drive to the relevant hunger drive, are experiments by Miller (6), Webb (10), and Amsel (1), which have shown how response strength may be increased through the introduction of additional irrelevant needs. Of these, Miller's experiment affords the best comparison with those of Estes and Skinner and of Mowrer and Viek, since it also employed an emotionally derived need as the irrelevant part of the motivational complex. In this study, rats were trained in a T-maze under relevant hunger motivation. When they had learned the correct response to food, they were satiated and half of their number was shocked briefly on a grid *separate from the maze*, immediately before running the maze. The animals which were shocked ran more quickly and with fewer errors. In this case, the irrelevant need of emotionality was facilitating to the previously learned response.[3]

Apparently the addition of an emotionally derived need to a motivational complex may increase or decrease the strength of a response learned prior to the presence of that need. The position to be taken here is that such seemingly conflicting results can be shown to depend upon the relationship of the drive stimulus (S_D), characteristic of the emotionally derived need, to the response in question. If the introduction of an emotionally derived need into an experimental situation provides a drive stimulus which has been conditioned to a response which competes with the one being observed in that situation, then, although the generalized drive level may actually be stronger, the frequency and magnitude of the original response might still be reduced. In other words, the introduction of the new need state, with its characteristic stimulus component, may be thought of as changing the response hierarchy in the given experimental situation. Such an analysis seems particularly relevant to the experiment of Estes and Skinner. A need to reduce anxiety, conditioned to the tonal stimulus, would be aroused whenever that tone was sounded in the Skinner box situation. The total motivational level might now be presumed to have increased by the addition of anxiety motivation to the dominant hunger motivation. However, other responses conditioned to

emotionally derived needs will be differentiated: *pain* (actually the need for pain reduction) which is contiguous with noxious stimulation; *anxiety* (or fear), which is the conditioned pain reaction; and *emotionality* (a term employed by Siegel and Stuckey, 9) which is a perseverative response of the organism following pain, thought of as having motivational properties. Pain and emotionality have the status of primary needs, since they are independent of learning; anxiety (fear) has the status of a secondary (learned) need.

[3] The experiment was repeated with shock administered for one sec. at the starting point of the maze and the results were the same.

the anxiety stimulation (crouching, escape responses, anticipatory responses to shock) might now successfully compete with the previously dominant lever-pressing response. It might appear, then, as if the addition of a need state to a motivational complex reduced the level of motivation.

If, on the other hand, the introduction of an emotionally derived need does not provide stimuli which give rise to responses which compete with the previously dominant one, then that response should be observed to increase in strength when the new need is introduced. This would seem to fit Miller's experiment. In that instance only one shock, of one sec. duration, was administered *away from the maze choice point.* It could not then be expected that the subsequent stimulation from emotionality would be conditioned to responses in the maze which could compete with the dominant maze-running tendency. Here, then, emotionality would serve only to raise the general motivational level, and hence the maze-running performance.

If the above analysis of the effect of emotionally derived needs upon previously learned instrumental responses carries over to consummatory responses, the following experimental prediction can be made: that animals trained to drink in specific drinking cages for a given time each day, and trained to apparent maximum performance under such conditions, will show an increased rate of drinking when the drinking period is immediately preceded by noxious stimulation administered in a situation other than the one in which the drinking occurs.[4]

[4] An experiment by Siegel and Stuckey (9), reported after the present data had been collected, is very similar to the present in its experimental aspects. The hypothetical basis and interpretation are quite different. These differences will be treated in the discussion section.

Another experiment to be reported (2), tests the prediction that, under the drinking conditions outlined, the rate of drinking will *decrease* when the introduced emotionally derived need (anxiety) has been conditioned *in the drinking situation.*

Experimental Procedure

Subjects and apparatus.—The subjects for this experiment were 20 female hooded rats from the colony maintained by the Psychology Department, State University of Iowa. Their ages ranged from 120 to 150 days at the start of the experiment.

The apparatus employed consisted of (1) a battery of individual cages in which the *S*s were confined for 10 min. each day for the measurement of water intake, and (2) a direct-current source connected to a grid for the delivery of electric shock.

The drinking cages were arranged in a row. The floor, top, and front and rear walls of the cages were of hardware cloth; the sidewall partitions were made of plywood painted dark grey to eliminate the possibility of any facilitating or interfering effect of the drinking of one *S* upon that of another as a result of visual stimulation. The inside dimensions of an individual cage were $8 \times 4 \times 6$ in.

Shock was administered in a compartment of dimensions $7\frac{1}{2} \times 4 \times 3\frac{1}{2}$ in., which was fitted with a grid floor and plexiglass top. The walls of the shock compartment were of three-quarter in. pine, painted white. The source of electric current was a 6c6 pentode vacuum tube provided with a plate voltage of 650 volts. An intensity of 50 microamperes was employed throughout as the noxious stimulus. The nature of the shock circuit was such that variations in bodily resistance of the animals produced no measurable change in current.

Habituation period.—During this 10-day period the *S*s were handled 10 min. per day by the experimenters. A drive rhythm based upon 22 hr. of food deprivation and 12 hr. of water deprivation was established.

Control period.—This was a 14-day period following habituation training during which daily measurements were taken of the amount of water intake for each *S* during a 10-min. period. Drinking performance was observed in the morning following 12 hours of water deprivation. Forty-five min. prior to drinking, each *S* was given four gm. of Purina dog chow in an individual feeding cage. Water was not available at this feeding, so that a further increase in thirst presumably occurred. Approximately one-half

hour after the cessation of eating, the *S*s were put into the drinking cages. The water bottles were then weighed and inserted into the cages. After a period of 10 min., the water bottles were removed and re-weighed. The *S*s were then fed the remainder of their 10-gm. diet and returned to the home cages. When 14 days of this kind of training had been carried out, the group drinking curve appeared to have reached a relatively stable asymptote. The experimental manipulation was, therefore, introduced at this point.

Experimental period.—During this three-day period the drive schedule was continued as before; however, a series of shock trials was given on each day between the prefeeding procedure and the measurement of water intake. Each *S*, at the completion of the four-gm. ration, was given four shock experiences. Each shock trial was of 20 sec. duration at an intensity of 50 microamperes. The interval between shock trials was approximately 10 min. On each day of the experimental period, the *S*s were placed in the individual drinking cages directly following the last shock experience and the water intake was measured as previously described. The remainder of the procedure remained identical with that during the control period. This procedure was followed for three days. Temperature readings were made on every day during the control and experimental periods. The small day to day temperature variations were found to be not significantly related to the average water intake over the relatively stable portion of the drinking curve.

Switched period.—This period followed immediately the experimental period and consisted of two additional days of drinking observation. On the first day of the switched condition the odd-numbered *S*s were shocked prior to drinking, as during the experimental period; the even-numbered *S*s were not shocked, as during the control period. On the second day of this period the procedure was reversed: the even-numbered *S*s were shocked while the odd-numbered *S*s were not. This was designed to serve as an additional check on the effect of shock-produced emotionality on drinking.

Results and Discussion

Comparison of control and experimental periods

The major findings are presented graphically in Fig. 1. This shows the mean water intake in grams of 20 subjects during the 14-day control period and the three-day experimental period. The curve of drinking, which had reached a relatively stable limit at about 5.7 gm. shows a sharp increase during the experimental period to a value of over seven gm. The rise of the control period drinking curve over the first nine days is probably due to the fact that on Day 1, the drinking situation was new to all *S*s. It probably mirrors the effects of two general factors: (1) adaptation of exploratory behavior which would compete with the drinking response, and (2) conditioning of drinking to the stimuli in the new drinking situation.

In order to make a statistical test of the difference shown between the control and the experimental periods, it was necessary to have a single score representing the drinking level of each *S* at the end of the control period and another single score for each *S* during the experimental period. To this end, a mean drinking score was computed for each *S* for the last three days of the control period, and another mean score was found for the three days of the experimental period. Table I shows these scores and the differences between pre-shock and post-shock drinking for each of the 20 *S*s. Examination of this table will reveal that in only two out of 20 cases was the pre-shock score higher than the post-shock score. The mean pre-shock drinking for 20 *S*s was 5.78 gm.; the mean post-shock drinking was 6.88 gm. An analysis of the difference for related measures yielded a *t*-value of 3.42 (for 19 degrees of freedom, the *t*-value at the one percent level of confidence is 2.86). A reasonably confident statement may, therefore, be made that the shock administered just prior to drinking resulted in a real increase in water intake. This result is in line with the hypothesis as stated.

An experiment recently reported by

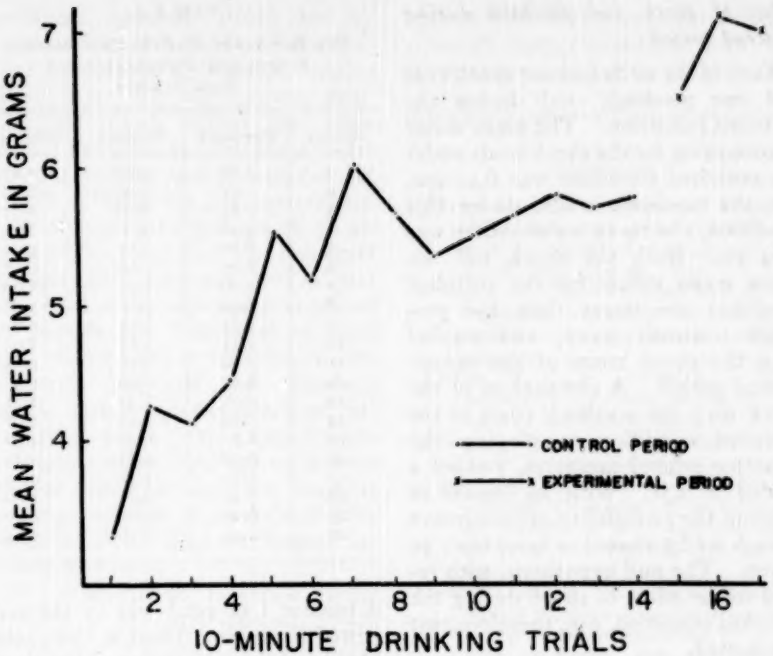


FIG. 1. Mean water intake in gm. of 20 *S*s on successive, daily, 10-min. drinking trials, under control and experimental conditions

Siegel and Stuckey (9) tested certain physiological hypotheses regarding thirst and reported experimental findings which are essentially the same as the present. They concluded that emotionality induces a state of thirst in the rat. In view of the fact that 27 out of 40 *S*s in their experiment did not drink *at all* over a two-hour period following faradic stimulation, such a conclusion may be seriously questioned. Of the *S*s that did drink, in the Siegel and Stuckey experiment, those which had previously been shocked drank more than the control animals, which had not been shocked. This would seem to suggest, as has been hypothesized here, that emotionality increases not the thirst level but rather the general motivational level, which would, according to the introductory formulation, result in an increase in the strength of *any* concomitant response. Since, in the case of the present experiment, and that of Siegel and Stuckey, the drinking response was the aspect of the total response pattern which was measured, this response was found to have increased. One test of these somewhat conflicting interpretations would involve the measurement of food intake following faradic stimulation. If, as Siegel and Stuckey suggest, emotionality results in a direct inhibitory effect upon salivary flow with resulting dryness of the buccal cavity, and not to a general unspecific increase in the motivational level, then the ingestion of dry food following faradic stimulation, where emotionality is *not conditioned to competing responses*, should be inhibited. According to the position taken here, emotionality would also serve to increase the rate and amount of food ingested under conditions similar to those of the present experiment.

Effect of shock and no-shock during switched period

Each of the 20 *S*s had one shock trial and one no-shock trial during the switched condition. The mean water consumption for the shock trials under the switched condition was 6.42 gm. For the no-shock trials, under this condition, the mean water intake was 6.09 gm. Both the shock and no-shock mean values for the switched condition are larger than the pre-shock (control) mean, and smaller than the shock mean of the experimental period. A comparison of the shock with the no-shock trials of the switched condition, employing the *t*-test for related measures, yielded a *t*-value of 0.71. With 19 degrees of freedom the probability of occurrence of such a *t* by chance is more than 40 in 100. The null hypothesis, with regard to the effect of shock during the switched condition, can, therefore, not be rejected.

The failure of a significant difference to be established between the shock and no-shock trials of the switched condition leads to some interesting conjecture regarding the conditionability of the response characteristics of emotionality to cues contiguous with them. If, during the experimental period, the perseverative emotionality from shock could be conditioned to the cues in the drinking situation, then, in the absence of shock before drinking, the drinking cage itself might elicit a conditioned form of the response of emotionality. If emotionality had previously been reduced by drinking, there might result an increase in water intake during the no-shock trials of the switched condition over what it was during the control period. Although the mean difference between pre-shock and no-shock (switched) was not found to be statistically significant ($t = 1.05$), the difference (.32 gm.) was in the suggested direction. There is, then, some likelihood that the failure of the shock and no-shock trials of the switched condition to show a significant difference in water intake may be attributed to this factor of conditioned emotionality during the no-shock, switched condition trials.

TABLE I

PRE-SHOCK AND POST-SHOCK DRINKING SCORES AND DIFFERENCES FOR EACH SUBJECT

Subject	Post-shock	Pre-shock	Difference
1	7.00	5.03	1.97
2	5.77	5.96	−0.29
3	7.83	6.50	1.33
4	5.90	4.66	1.24
5	6.76	5.83	0.93
6	8.03	7.36	0.67
7	4.00	5.00	−1.00
8	6.36	5.20	1.16
9	5.13	4.53	0.60
10	7.73	5.83	1.90
11	5.80	5.56	0.24
12	8.53	7.96	0.57
13	8.96	6.33	2.63
14	7.60	5.53	2.07
15	8.40	6.90	1.50
16	6.56	5.23	1.33
17	3.60	3.30	0.30
18	7.30	5.50	1.80
19	6.96	5.83	1.13
20	9.50	7.36	2.14

A more definite answer to the question, can emotionality (the perseverative emotional response from painful stimulation) become conditioned to contiguous stimuli?, must await further experimentation, wherein a greater number of trials are allowed for this conditioning to occur. In the present experiment, the emotionality response could only have been conditioned to the cues in the drinking situation during the three days of the experimental period.

SUMMARY

In order to test the hypothesis that the magnitude of a consummatory

(drinking) response would be increased when that response was preceded by noxious stimulation administered apart from the drinking situation, 20 female albino rats were trained to drink in individual cages during a 10-min. period each day for 14 days. When the average amount of water consumed had ceased to increase from day to day, and remained approximately constant for several days, a three-day experimental period was introduced. During these three days, electric shock stimulation immediately preceded the drinking period. After the experimental period, there was a further two-day period during which 10 of the 20 Ss were shocked before drinking on the first day, and the other 10 were shocked on the second day. This was termed the switched period.

The introduction of shock before drinking in the experimental period resulted in a sharp and significant rise in the curve of average water consumption. On the basis of these results, the hypothesis as stated could not be rejected. The results were interpreted in terms of an increase in the total generalized drive strength, and this interpretation was compared with an alternative interpretation of similar results by other experimenters.

The data from the switched condition did not show a statistically significant difference between the amount of water consumption with and without prior shock experience. These results were interpreted as indicating that emotionality (the perseverative aspects of the pain state) may become conditioned to stimuli which are contiguous with that perseverative state although not contiguous with the shock experience itself.

(Manuscript received September 2, 1949)

References

1. Amsel, A. The combination of a primary appetitional need with primary and secondary emotionally-derived needs. *J. exp. Psychol.*, 1950, **40**, 1–14.
2. Amsel, A. The effect upon level of consummatory response of the addition of anxiety to a motivational complex. (In prep.)
3. Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, **29**, 390–400.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century Co., 1943.
5. Kendler, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.
6. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization, *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
7. Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193–200.
8. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
9. Siegel, P. S., & Stuckey, Helen S. The effect of emotionality on the water intake of the rat. *J. comp. physiol. Psychol.*, 1949, **42**, 12–16.
10. Webb, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.
11. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.

RESPONSE LATENCY AT ZERO DRIVE AFTER VARYING NUMBERS OF REINFORCEMENTS [1]

BY DAVID ZEAMAN AND BETTY J. HOUSE

University of Connecticut

Introduction

The present experiment was designed to test a specific theory, but the data may be relevant to any theory which attempts to account for behavior in terms of previous learning and present motivation.

The hypothesis to be tested is Corollary II (12, p. 247) of Hull's Postulates 5, 6, and 7 which states:

Many experimental tests of this corollary can be made—at least as many as there are combinations of different measures of ${}_{S}E_{R}$ and drive. Some of these combinations have been tried by other investigators with contradictory results. The particular combination used in this study has been chosen, not only because it was previously uninvestigated, but also because it obviates certain difficulties reported with others.

A descriptive summary of the relevant experimentation cannot be given here. The literature is large (2, 4, 7, 9, 11, 15, 17, 19, 21, 23, 24, 25, 26, 29) and much of it current.

A critical summary of the experimental evidence leads to the conclusion that a number of difficulties are inherent in the type of research designed to discover the strength of apparently unmotivated behavior.

The first of these is that of finding a set of empirical operations which may be said by convention to create zero drive, when the drive is to be regarded as an inferred state of the organism conceptualized as a theoretical construct 'drive' or *D*. That there is no adequate convention as yet is attested to by the wide variety of elaborate satiation procedures previously used to reduce the primary appetitive drives to zero. Zero level for some of the primary aversive drives such as the light aversion drive in the present research provide an easier problem, but there has been little experimentation to take advantage of this fact.

The second difficulty is similar to the first. It is that of assigning to some empirical observation of behavior the status of anchor for zero amount of response strength which may be conceptualized, as in Hull's system, reaction potential or ${}_{S}E_{R}$. If a theory asserts that there will be strength of behavior under zero drive and at the same time does not establish explicit operational definitions which will tie response strength and zero drive to some operations or observations, then no clear-cut relationship exists between theory and data.

The difficulties are not insurmountable. It is necessary only to try out various conventions until some are hit upon which bring about the happiest relation between theory and data. Many existing conventions relating operations of satiation to the magnitude of resultant drive have not proven to be the best of choices. They do not make clear the grounds for deciding which of any two satiation methods produces the lower degree of drive. Why, for example, does interrupted access to wet mash produce lower hunger drive than continuous access to dry pellets? Because one method gets more food into the animal? Or because one produces lower resultant response strengths? Both or either of these grounds may be implied when experimenters describe their satiation methods as 'more rigorous' or 'more ex-

[1] This experiment was performed in the psychological laboratory of Brown University.

When primary drive strength (*D*) is zero, reaction-potential (${}_{S}E_{R}$) has an appreciable but relatively low positive value which is a positive growth function of the number of reinforcements.

treme,' but they are not explicated. Explicit adoption of either of these grounds does not completely solve the problem. If we scale D by the amount of food a given procedure can get into the animal, what amount shall serve to indicate zero drive? The rigorous satiation procedures used by Koch and Daniel (17) and by Webb (26) might be offered as convention to be given the status of zero hunger drive. However, it may be possible through other extreme methods, which might make additional use of social facilitation (10), specific appetites (28), or punishment of not-eating behavior, to get even more food into the animal. In these cases, what value of drive is to be assigned to these operations? It is relevant in the context of the present study on light drive to point out here that such a problem does not arise in conjunction with some aversive drives. If we identify the strength of a drive with the strength of some external stimulus, we may offer as convention a zero amount of this stimulus as corresponding to zero drive strength. For example, with an aversive light drive, no quanta of light would correspond to no drive at all. It is possible that people can go on finding more and more ingenious ways of further food-satiating an animal, but it is unlikely that anyone will wish to go further than no quanta of light at all as a measure of zero light drive.

Returning now to a consideration of the second implicit ground for anchoring D mentioned above, that of scaling D by the amount of response strength produced by a given satiation procedure: if one feeding method is stated to be more rigorous than another because it depresses behavior further, it is implied that zero drive would be associated with that method entirely eliminating behavior. Such a convention would, of course, make the relation between zero drive and response strength (${}_{S}E_{R}$) a nonempirical matter. A definition of zero drive as one producing zero ${}_{S}E_{R}$ would be logically inconsistent with Hull's theory which postulates that zero drive will not entirely eliminate behavior strength, hence it could not be used very well in an attempt to test Hull's theory.

This does not mean that an aspect of behavior cannot be used to infer the strength of a drive produced by some experimental operation. Circularity will be avoided simply by inferring zero drive from an aspect of behavior other than the one under investigation. The drive state bears many relations to behavior. The presence of it not only maintains behavior (the ${}_{S}H_{R} \times D$ relationship) but the reduction of it can strengthen behavior under appropriate (reinforcing) circumstances. One might offer as convention for zero drive that level of drive the reduction of which produces no measurable learning under otherwise optimal conditions. The experimental operation associated with such a level can then be carried over to already existing, relevant habits of various strengths to find out whether zero level of drive so defined would reduce performance to zero level.

A practical illustration of this method applicable to light drive may make the notion clearer. A rat is put into a cage wired in such a way that his running to one side of the cage turns out an overhead light. If the light is very bright, he will learn to prefer the side which brings about the dark. Different rats may be tried under various lesser degrees of illumination until one is found too dim to produce learning. In this way, a threshold of reinforcement by light drive reduction is found. Such an experiment has already been performed by Hanson (9), who found a reinforcement threshold between .135 and .05 millilamberts. The rats spent equal time in the light and dark at this level of illumination. Now if this threshold level (or less) of light intensity is carried over into the present experiment after varying amounts of training have already taken place under strong drive, the question of the joint effect of zero drive (response defined) and varying habit strengths on performance may be checked. Similar techniques are conceivably employable with any primary drive, appetitive or aversive.

Establishment of an adequate zero

drive convention was the first difficulty seen in the literature on the present problem. Of equal importance is the problem of assigning the value of zero response strength or reaction potential to some empirical observations. In the literature cited above, the following measures are some that have been interpreted as indices of zero strength: chance performance on a maze, the level of strength produced by extinction, criterion latencies, zero number of responses to an extinction criterion, a 'negative' number of extinction responses, and the unconditioned level of strength. With so large a number of 'zero' points, contradictory results are to be expected. In the interests of parsimony and consistency some one of these should be elevated to the status of accepted convention. The last of these conventions, the unconditioned strength —what Skinner calls the free operant level—has the most to recommend it. It is the most general convention in that it will yield a 'zero' level for many measures of response strength such as probability, latency, frequency, amplitude and so on; and it ties down 'zero' level of reaction potential to independent measures of behavior. Estimates of the unconditioned strength can be determined empirically for both the Pavlovian and selective type learning situations.

The operant level criterion of zero reaction potential, as stated, is ambiguous in that two different methods of measurement are possible. As a first possibility, some single set of experimental conditions could be set down as the standard conditions under which operant level should be measured. For example, the untrained, unacclimatized animal could be put into the experimental situation under maximum drive for a given period of time or number of sessions to allow sampling of his unconditioned 'zero' strength. One value of unconditioned strength for each response measure would be obtained by this method. The second possibility would require the measurement of operant level under many conditions and would undoubtedly result in a shifting 'zero' point. In a given experiment, the effects on operant level of any experimental operations or conditions other than those involving reinforcement would have to be determined. Changes in the strength of relevant or irrelevant drives, for example, may operate to change unconditioned response level. Also, continued time or trials in the situation without reinforcement before learning may result in either an increment in unconditioned strength, as Mote and Finger (18) have demonstrated for a running response, or in a decrement, as shown by Schoenfeld, Antonitis, and Bersh (22) for a lever-pressing response. The second convention of a shifting zero point is the more appropriate for a reinforcement theory on the grounds that it partials out of the total variation in performance changes in strength not attributable to the usual, independent theoretical variables. Age, sex, and heredity are examples of variables that as yet play no formal part in reinforcement theory. Their effects are usually partialled out statistically or experimentally. Changes in unconditioned strength belong in this category of as-yet-not-theoretical variables and should therefore also be cancelled out. Adoption of the shifting zero point would provide a control for this factor. As an example illustrating the need for such a control, a recent experiment (26) may be cited in which various strengths of irrelevant drive served as the independent variable. The resulting increments in performance strength were attributed to the energizing effects of the irrelevant drive on the strong habit structures that were present. Whether or not similar increments in performance strength would have occurred in the absence of habit structure was not determined. A necessary control for this interpretation would have been the measurement of operant level under each of the irrelevant drive conditions and the determination of the difference between the conditioned and unconditioned levels. Use of this difference as a measure of reaction potential would, of course, amount to adoption of the shifting zero point convention.

A difficulty in measuring operant level for a zero point may appear in those cases

where the conditioned response before learning is not sufficiently strong in the presence of the conditioned stimuli for any immediate aspect of it to be measured. Under these circumstances, some investigators have measured the unconditioned strength by the number of reinforcements required to bring some aspect of the response past its threshold into the experiment. Measuring unconditioned level from the negative or below-threshold strength is not difficult in the Pavlovian set-up because reinforcements can be counted by stimulus pairings before the CR appears, but a problem arises in operant conditioning where the reinforcements are contingent upon the presence of the response from the outset. How can a negative unconditioned strength be computed here? Although the rat usually comes into the typical operant (lever-pressing) situation with a low measurable rate of response, some investigators arrive at a below-threshold strength in the following way. If the CR fails to occur within a given period of time after the presentation of the CS (the lever), the response is forced by baiting the manipulandum with food. The attempt to get at the bait brings about the response, which is then reinforced with a food pellet and counted as an 'unconditioned' reaction (5). The number of such reinforcements required to bring the response within the set time limit measures the amount of below threshold strength and hence the distance from no responses at all down to the 'true' zero of strength. As an independent empirical support for this measure of negative strength, Hull (12, p. 325, 8) offers the negative strength arrived at by extrapolation of certain learning curves to the region of zero number of reinforcements.[1] It might be recommended here that if a time limit is used in a criterion of presence or absence of response in a measurement of operant level before learning, the same time limit be used in resistance-to-extinction criteria. This is not generally the case. Otherwise, this method can be considered as an adequate one of assessing the free operant level to conventionally represent zero reaction potential.

In summary, the difficulties apparent in this area of research are primarily those associated with the absence of established conventions for zero drive and zero reaction potential. For such a problem there is no solution other than the continued use by investigators of explicit conventions which are general, consistent with theory, productive of unequivocal results, and technically feasible.

The two conventions respected in the present research are a response-inferred zero point for drive and the unconditioned level of strength as zero point for reaction potential. Explicating more fully, zero drive is that level of drive the reduction of which cannot produce learning (increase in reaction potential) in an otherwise optimal reinforcement situation; and zero reaction potential is the positive or negative strength of performance measured before reinforcement has begun.

It is the purpose of the present research to test the part of Hull's theory stated at the outset, using these two conventions.

Procedure and Apparatus

Subjects.—The subjects were 38 albino rats from the colony of the Brown laboratory, 21 males and 17 females. Their ages ranged from

[1] Full recognition is apparently not always given to the fact that Hull explicitly adopts this convention allowing zero $_sE_R$ to be anchored to a negative response strength. When, for example, Koch and Daniel (17) found a median of one extinction response at 'zero' drive, they regarded this strength as strikingly close to zero reaction potential and hence contrary to Hull's theoretical prediction of a greater than zero reaction potential at zero drive. Their assumption was that no-extinction-responses-within-five-min. was equal to zero $_sE_R$, in contrast to Hull's convention which would make zero extinction responses equal to greater than zero $_sE_R$. Hence, although Koch and Daniel's data prove that the Hullian equations overestimate response strength at 'zero' drive, they do not controvert the part of the theory that predicts greater than zero $_sE_R$ at zero drive. The reason, as has been pointed out, lies in this particular convention relating zero $_sE_R$ to behavior.

100 to 150 days at the beginning of the experiment. Half of the rats had been used previously in a panel-pushing experiment involving a food reward. As a group, their performance was not distinguishable from that of the naive rats. The differences in sex and previous training were randomized among the experimental conditions.

Apparatus.—Essentially, the apparatus consisted of a two-compartment box, with a swinging door which allowed the rat to go from one compartment to the other. Between trials the swinging door could be enclosed by a double sliding panel. A light source was so arranged that either compartment could be darkened while the other remained illuminated, or both compartments could be made light or dark simultaneously.

The principal features of the apparatus are shown in Fig. 1. The lower part of the drawing shows the experimental box, with dimensions eight in. wide, two ft. long, and eight in. high. The partition across the center of the box divides it into two compartments $8 \times 12 \times 8$ in. A doorway, two in. wide and three in. high, was placed in the middle of the partition. A thin piece of metal of the same dimensions as the doorway swung freely so that a rat could easily push his way from one compartment to the other. A double wooden panel covered the swinging door between trials. This panel moved in two sets of grooves and could be raised or lowered by the experimenter by means of a string. When lowered, it enclosed the swinging door so that between trials the rat in either compartment was presented with a smooth wooden panel over the door. The animals were placed in the apparatus and removed through the two doors in the ends of the box. The box was constructed of Masonite with a hardware-cloth floor. Set into the front of the box was a piece of glass 20×7 in. which allowed the experimenter to observe the rat's behavior. The top of the box was covered by opal flash glass. The inside of the box was painted white throughout, except for the swinging door, which was painted black. The box was painted flat black on the outside.

The upper part of Fig. 1 shows the wooden framework supporting the light source. Two 300-watt bulbs with built-in reflectors were centered over the two compartments of the box. Below each of the bulbs was a Pyrex cake plate, $8 \times 11 \times 2\frac{1}{2}$ in., filled with water to absorb the intense heat. Next in sequence was a sliding filter holder. For the present experiment, the filter holder contained a sheet of aluminum metal foil. In the figure, the sheet of metal paper

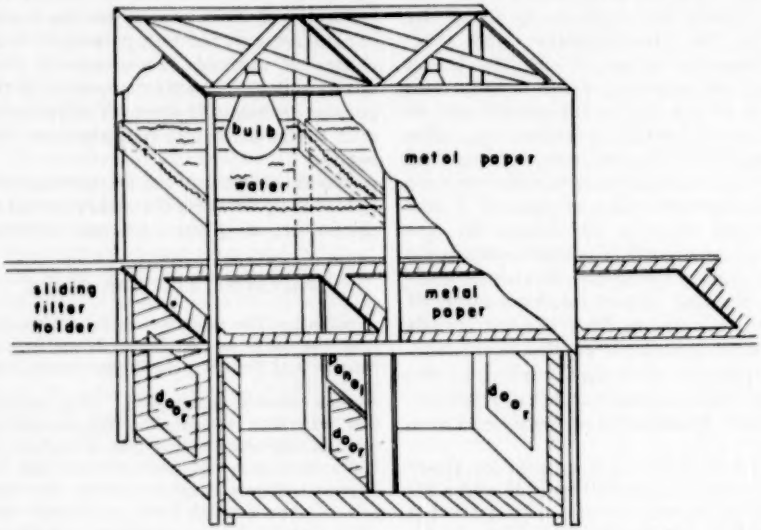


FIG. 1. The experimental apparatus described in the text. When the light is on in the rat's compartment, he runs through the door into the dark compartment as soon as the panel is raised. The black cloth and metal paper have been torn away to show the inside of the box. Not shown are the two layers of light-diffusing glass directly on top of the compartments, the pulley arrangements for the filter holder and panel, and the electrical wiring.

covers the right compartment and cuts off the light leaving the compartment dark, while the left compartment is illuminated. The experimenter, by pulling a string attached to the left end of the filter holder, could slide the sheet of metal paper over the left compartment, leaving the right side illuminated. Directly beneath the filter holder was a sheet of frosted glass. The frosted glass, together with the opal flash glass covering the compartments, acted to diffuse the light from the two bulbs. The white walls of the inside of the box provided an environment of relatively constant illumination. The brightness of a lighted compartment was 743 millilamberts, as measured by the Macbeth Illuminometer. A dark compartment, shielded by a layer of metal paper, had a brightness of .007 millilamberts. The wooden framework containing the light source was covered on three sides by aluminum metal foil painted flat black on the outside to radiate as much heat as possible. The front of the structure was draped with several layers of black-out cloth enabling easy access to the water pans for frequent re-filling. Enough of the bulb-heat was thus radiated off and absorbed to keep the temperature inside the two compartments within one degree of that of the room over a period of five hours.

Recording system.—The apparatus was wired for electrical recording of latencies. A microswitch on each side of the box could be closed or opened by movement of the floor of the box which tilted slightly under pressure. Thus, if a rat were on the right side, the right floor-switch was closed, and the left floor-switch was open. Either of these two switches could be placed in series with a Springfield timer by means of a single-pole double-throw switch. For example, if the rat were on the right side of the box, *E* could prime a circuit joining the timer with the right floor-switch by throwing the single-pole double-throw switch to the right. The circuit thus formed would be closed if the rat were on the right side of the box, and if the sliding panel were raised. Raising of the panel also gave the rat access to the swinging door and constituted the beginning of a trial. When the rat pushed open the swinging door and went into the next compartment, the right floor-switch was opened, breaking the circuit and stopping the clock. Since the closing and opening of the circuit coincided with the beginning and end of a trial, the latency of response could be recorded from the timer.

Learning—under drive.—Learning trials were conducted in the following manner. A rat was placed in a dark compartment. A few sec. were allowed for the experimenter to close the door of the compartment and to make sure that the switches were set correctly. When *E* was ready to begin a trial, he pulled the filter holder, moving the sheet of metal foil from one compartment to the other, thereby illuminating the compartment containing the rat. Two sec. later, *E* raised the sliding panel, exposing the black swinging door and starting the Springfield timer. If the rat pushed his way into the dark side of the box, the floor-switch was opened and the clock stopped. If the rat made no response within five min. after the raising of the panel, he was gently pushed through the swinging door and his latency recorded as indeterminate. In either case, *E* lowered the sliding panel and allowed the rat 60 sec. reinforcement in the dark compartment. During this period, he recorded the latency and re-set the clock to zero. After 60 sec., the single-pole double-throw switch was thrown to the side containing the rat, and the filter holder was pulled to the other side so that the rat was now in the light. Two seconds later *E* raised the sliding panel to begin the next trial and to start the clock.

The learning trials were given in daily sessions of 10 trials. Five experimental groups were brought to different levels of learning by varying the number of learning trials. Six rats were given 10 learning trials, 15 rats had 20 trials, eight rats had 40 trials, four rats had 70 trials, and five rats had 100 trials. The group containing the 15 rats was run as part of another experiment requiring identical conditions. For the other groups in the present experiment, smaller numbers of rats were found to give as regular results.

No drive—after learning.—On the day following the learning trials, each experimental group was given 10 test trials to determine strength of response under no drive. The same procedure was followed for test trials as for learning trials, with the obvious difference that during these trials both of the compartments were dark. Thus the 60-sec. period between trials did not serve as a primary reinforcement. Response strength under no drive for the five experimental groups was determined after 10, 20, 40, 70, and 100 learning trials, respectively.

No drive—before learning.—As a control, eight untrained rats were given 10 trials a day for 10 days with both compartments dark. These trials were conducted with exactly the same procedure as the test trials described under the section labelled '*No drive—after learning.*' In this way, there was determined the unconditioned operant level of response under zero drive and the effect on this level of continued 'elicitation' without reinforcement.

Drive—before learning.—As an additional control the unconditioned operant level under the intensity of drive used in the experiment was determined. Six of the same rats used in the

TABLE I

PERFORMANCE EXPRESSED IN LATENCIES UNDER THE FOUR DIFFERENT EXPERIMENTAL CONDITIONS

The medians are the central tendencies of groups of 10*N* values.

Daily Block of 10 Trials	Learning Under Drive		No Drive After Learning		No Drive Before Learning		Drive Before Learning	
	N (Rats)	Log Mdn. Latency	*N* (Rats)	Log Mdn. Latency	*N* (Rats)	Log Mdn. Latency	*N* (Rats)	Log Mdn. Latency
1	38	1.59	6	1.72*	8	2.28	6	2.45
2	32	1.02	15	1.53	8	2.15	6	2.48**
3	17	.93			8	1.85	6	2.28
4	17	.95	8	1.34	8	2.26	6	2.35
5	9	.74			8	2.31		—***
6	9	.65			8	2.05	6	2.36
7	9	.65	4	1.22	8	2.10	6	2.48**
8	5	.48			8	1.88	6	2.48**
9	5	.65			8	2.20		
10	5	.48	5	1.00	8	2.13		

* For the values in this column, the number of the daily block of 10 trials refers to the number of *previous blocks of learning trials under drive.*

** Criterion values representing 'greater-than-five-min.'

*** This group of rats was run at an unscheduled hour on this day. The data were discarded.

first control group were given 10 trials a day for seven days with both of the compartments lighted. These trials were given two weeks after the last of the 'No drive—before learning' trials.

RESULTS

The major results of this investigation are given in Table I and reproduced graphically in Fig. 2.

Learning—under drive.—The curve labelled 'Learning—Under Drive' represents the course of latency change produced over 100 trials when a light intensity of 743 millilamberts provided the drive and a reduction of this light to a level of .007 millilamberts provided the reinforcing situation (or reinforcement). The negatively accelerated decreasing shape of the curve is that usually found in learning curves when latency is the measure of response magnitude. The amount of change, one log unit, is relatively small, however, and may be attributed to either the relative massing of trials or the weakness of the light drive.[3] The major part of the latency change occurs within the first 50 trials. The last 40 latencies are relatively asymptotic. There is a considerable amount of variability in this function despite the fact that each point represents the median of at least 50 values.

No drive—after learning.—The curve labelled 'No Drive—After Learning' supplies a confirmation of the hypothesis which this investigation sought to test, that under no drive, performance would be a positive growth function of the number of reinforcements. If one assumes an inverse relationship between latency and strength of performance, the confirmation of the hypothesis is clearly suggested.

It is important to note that for this function the abscissa refers to the *number of previous blocks of learning*

[3] Experiments by Keller (15) and Hanson (9) found behavioral maxima at much lower intensities of light drive. Their experiments will be discussed in a later section.

trials under drive. The function was plotted in this way so that the vertical distance between it and the learning curve would represent the change in latency at each stage of learning produced by the removal of drive. The log median latency at each of the five stages of learning is lower than that preceding it. There are no inversions of the points. The relatively smooth and progressive decrement of the five points is perhaps the strongest argument that can be made for there being a true function. A Festinger *d*-test (6) shows that the two extreme groups are significantly different between the .05 and .01 level. The large and consistent difference between the 'No Drive—After Learning' curve and the 'Learning—Under Drive' curve makes statistical test of the difference superfluous. Inspection of the 'No Drive—After Learning' curve might be all that is required to show that Hull's hypothesis would predict both the position and form of the curve drawn through the experimental points. The details of the relation between theory and data will be given in the Discussion.

No drive—before learning.—The separate group of animals who ran from dark compartment to dark compartment had their latencies undergo change described by the curve labelled 'No Drive—Before Learning.' There is essentially no systematic change in these latencies. The best-fitting straight line as determined by the method of averages shows a slight negative slope. The slope constant of the equation was −.02, the intercept

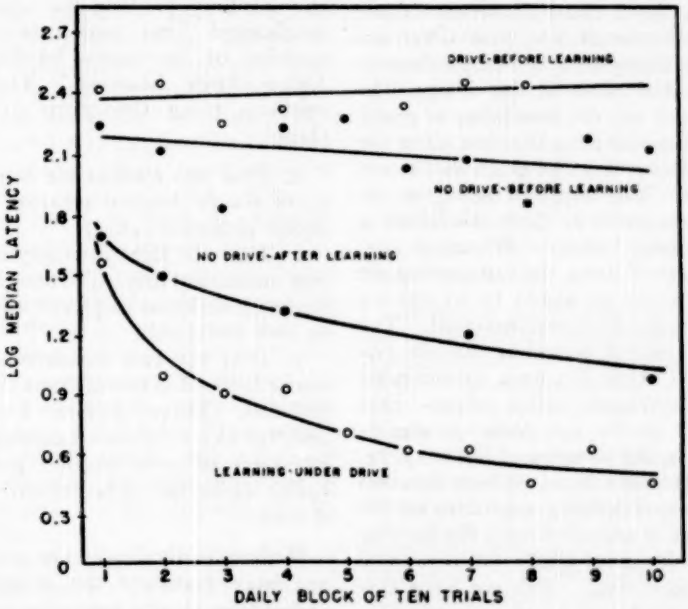


FIG. 2. A graph of the data of Table I. The two topmost functions represent the operant level under drive and no drive. The lowest set of points is a learning curve and the function on top of it shows the effect of removal of drive at varying stages of learning. With the exception of one, all points represent the median of at least 50 measurements.

2.23 log sec. The amount of decrement produced by this negative slope over the 100 is slight compared to that produced by learning.

Drive—before learning.—Serving as another control group, those animals run from lighted compartment to lighted compartment yielded latencies which varied in a manner shown by the curve marked 'Drive—Before Learning.' Again, no systematic change is evident during 80 trials. A straight line fitted to these data shows a slightly rising slope and an ordinal position above that of the 'No Drive—Before Learning' curve. The effect of the light is to depress responding in this situation. The effect is significant as shown by the fact that all seven of the points representing operant level under drive are higher than the corresponding points representing operant level under no drive. There are two reasons why this difference may be interpreted as a true difference due to the effect of the drive. The first rules out the possibility of group differences between the rats, since the same group was run under both conditions. The effect of being in the situation under no drive conditions is to decrease latency. Hence, if any, the effect of using the same group for both conditions would be to operate against the effect demonstrated. The second reason considers outside evidence. There has been considerable experimentation which shows that activity in the rat decreases significantly in the presence of light (**13, 14, 20**). It might be noted here that one of the usual defining conditions for the presence of animal drive is the *increase* in activity level under the presumed drive condition. Here we have evidence to the contrary. A stimulus condition has been clearly demonstrated to be a drive condition in that its reduction acts as a reinforcement. This same drive condition has been shown to decrease activity when no reduction is brought about by the animal's responses. The decrease in activity, i.e., increase in latency, under drive is actually greater than that shown in Fig. 2, since three of the plotted points have the indeterminate value of 'greater than five minutes.' The straight line fitted to the points has little quantitative meaning other than a rough indication that the operant level under no drive is considerably higher than under drive.

Discussion of Results

A direct confirmation of Hull's Corollary II is to be seen in the curves of Fig. 2. Assuming the adoption of certain conventions relating theoretical and experimental terms, the corollary predicts the negatively accelerated form and intermediate position of the curve labelled 'No Drive—After Learning.' The conventions tying theory to data are these:

1. That the median log latency is some simple inverse measure of reaction potential ($_sE_R$);
2. That the light intensity is a direct measure of drive (D), and that the degree of darkness employed was equal to zero drive; and
3. That the two measurements of unconditioned or free operant strength (labelled 'Drive—Before Learning' and 'No Drive—Before Learning') are estimates of zero reaction potential ($_sE_R$) under two different drive conditions.

If these equivalences are accepted, one may translate the graph into verbal form closely approximating the corollary under test. That there is 'appreciable' strength without primary motivation is empirically estab-

lished by the large area between the 'No Drive' curves before and after learning. The extent of this area brings into question that part of the corollary which states that without motivation ${}_{S}E_{R}$ will have a low positive value. The strength is equal to approximately 50 percent of the maximum strength produced by learning under drive. Maximum strength, using our convention of zero ${}_{S}E_{R}$, is the area between the operant level under drive and the asymptote of the learning curve under drive. It should be recognized, however, that these area relationships are dependent upon the transformation used to convert raw latencies to reaction potential. Other transformations than the median log latency, such as reciprocal latency, log reciprocal latency, geometric mean, etc., radically alter the percentage of maximum obtained, and even the form of the relationships. Independent of such transformations is one estimate of the magnitude of strength under no drive after learning. This would involve the comparison between the number of trials required under conditions of drive and no drive to produce the same level of strength. For example, after 100 reinforcements, removal of the drive reduces strength of performance to a level equal to that produced by 20 trials of learning under drive. This comparison would yield a percentage of 20 percent, which might more easily be interpreted as a low value. These values of 50 and 20 percent give substance to that part of the corollary predicting a relatively low positive value under no drive.

The verbal content of the corollary is completed with the statement that under no drive "${}_{S}E_{R}$ . . . is a positive growth function of the number of reinforcements." The negatively accelerated shape of the 'No Drive—After Learning' curve may be derived from this statement, and hence supports it. Thus all parts of the corollary are seen to have some basis in fact. There remains the task of examining the acceptability of the conventions adopted, the relation of the results to previous findings, and alternative theoretical explanations.

The goodness of a set of internally consistent conventions is partially determined by the consequences of its adoption for a particular theory. The use of the log median latency as a simple inverse measure of reaction potential is justified by the large and growing number of studies which have added to the validity of many of Hull's postulates and corollaries by employment of this transformation. In the present study, not only was a particular corollary confirmed, but the negatively accelerated appearance of the log latency learning curve is consistent with Hull's postulates concerning the growth of habit strength. These quantitative properties would be undisturbed if the distances between the operant level and the post-learning performance curves were plotted, thereby converting the functions into direct rather than inverse measures of ${}_{S}E_{R}$.

The advantage of a latency measure in the investigation of the present problem is derived partly from the disadvantages of the resistance to extinction measure most frequently employed in previous zero drive experiments. The measurement of the number of extinction responses to a criterion requires a period of time during which changes in drive may take place. Kimble (16) and Horenstein (11) have shown that drive increases very rapidly after food-satiation as indicated by changes in latency and amount of food intake. A further disadvantage of resistance to extinction measures is that a limited number of measurements can be obtained from each animal. In

the present experiment, ten latencies were obtained from each animal under no drive. These are practical advantages serving to recommend latency as a proper measure to be used in testing the zero drive corollary.

The convention relating light intensity directly to drive (*D*) has precedence in the experiments of Keller (15) and Hanson (9). They found that progressive increases in light intensity had behavioral effects which would lead us to infer that *D* had undergone progressive increments. Keller reports that three different measures of strength of a lever pressing response were related to the intensity of light used as a drive. Two measures of frequency of response increased with negative acceleration as the light intensity increased from 17 to 514 millilamberts. Latency of response decreased with negative acceleration over the same range. The asymptotes of response strength occurred at approximately 340 millilamberts, an intensity well below the 743 millilamberts used in the present experiment. Hanson has found that in a two-choice light-dark situation rats averted a light intensity of 6.8 millilamberts 93 percent of the time in the situation. As the light intensity was gradually decreased the rats' light-dark choice approached the chance level. No preference between light and dark was found when the light intensity reached .05 millilamberts. This value of light intensity approximates zero drive according to our previously stated convention of zero drive as that incapable of producing learning. The level of darkness actually used in the present experiment (.007 millilamberts) was well below the reinforcement threshold and approached a stimulus specification of absolute zero drive, defined as no light at all. The adoption of these operational definitions relating drive and light intensity allows the logical incorporation of the data of the light drive experiments within Hull's theory and provides a method for a clear-cut test of the corollary under investigation.

Use of the light drive is perhaps more suited to this general problem area than the more widely used appetitive drives for two reasons. Since the problem concerns primary drive and primary reinforcement, the possibility of secondary reinforcement effects must be eliminated. The secondary reinforcing aspects of food and water rewards make this factor difficult of control. It is unlikely, however, that the animal will come into an experimental situation with any secondary reinforcing associations connected with the light drive. The difficulty of independent experimental control of hunger and thirst constitutes a second drawback of the appetitive drives. Deprivation of water may, for example, prevent complete satiation of food, but the light drive does not interact in a similar fashion with other drives.

The last convention unifying theory and data is the shifting zero point convention. The consequences of its adoption were elaborated in the Introduction of this paper.

Considering now, the relation of our results to those of previous studies, agreement is found with all those studies (2, 4, 7, 15, 24, 25, 26, 29) which have reported the presence of performance strength under zero drive. Our results disagree with those studies (17, 23) which did not. For the specific problem of strength under zero drive after varying numbers of reinforcements, Siegel's experiment (23) is the most closely related to the present study. Some of his results can be interpreted as consistent with those of this experiment, and some as contradictory. The experiment most similar to the present one in its use of aversive light drive is that of Keller (15). He also found that removal of drive after learning left performance stronger than operant level.

The majority of studies concerned with zero drive indicated a presence rather than an absence of behavior strength. For the most part, those reporting negative results have used appetitive drives, with extreme satiation procedures. These may involve an artifact. Providing conditions for the animals to 'overstuff' themselves could produce a negative drive state leading to 'resting' re-

sponses as a means of reduction. One consequence of such an inferred drive state would be the interference with the trained response otherwise observed at zero hunger drive. The kind of evidence needed to further support the 'stuffed-state' inference would be a sharp decline in operant level at some point along the satiation continuum and/or a sudden interference with strong habits motivated by other drives at a similar point in food-satiation.[4]

In interpreting the results of this study, it has been demonstrated that these confirm the verbal content of the corollary of Hull's theory under test. It is also true that nothing in the results would negate the postulated mechanism from which the corollary was derived. The strength under zero drive is postulated to reflect the sensitization of habit structure by irrelevant drives. As a matter of fact, the present research is one of the few which shows that habit structure is necessary for the demonstration of apparently unmotivated behavior. Whether or not the habit structure is energized by irrelevant drives is not tested by our experiment, although this could most easily be done with the light drive technique, uncomplicated as it is by physiological interaction with other drives.

Various other interpretations of the data could, of course, be made. The results are consistent with Anderson's notion of an externalized drive (3), with Allport's functional autonomy (1), and with Woodworth's idea of a mechanism as a drive (27). Explanations of the results may possibly be formulated in terms of secondary reinforcing or secondary drive stimuli, although such explanations would be restricted to response produced or internal stimuli only. This derives from the fact that the two compartments are nearly identical. In running from dark compartment to dark compartment, the animal neither gains nor reduces any external stimuli. The inference of kinaesthetic stimulation acting as a secondary reinforcement of the responses producing it would serve to account for these results and other like it.

Although the data might be interpreted as supporting any of the theories mentioned, the formulation of Hull appears to be the most explicit and the most general.

[4] The frequently heard bit of dinner table conversation, "I'm too stuffed to move," may be offered as a kind of anecdotal analogue of this inferred negative drive state.

Summary and Conclusions

A light aversion experiment has been performed with white rats to determine the functional relationship between the strength of performance under zero drive and the number of previous reinforcements under maximum drive. The strength of performance under each drive condition was measured by the difference between the median log latency of response after training and the unconditioned, or free operant level before training. The response whose latency was measured may be molarly specified as a 'cross-the-box-through-the-swinging-door' response. Two zero drive conventions were considered, one a response-definition of zero drive as that level incapable of producing learning, the other a stimulus specification of zero drive as the total absence of an aversive stimulus. The low level of illumination employed as zero drive met the first criterion and approached the second. The major findings of this study are enumerated below.

1. When the primary light drive was reduced to zero, the strength of performance was found to be an in-

creasing, negatively accelerated function of the number of previous reinforcements under light drive.

2. The performance strength under no drive was approximately 50 percent as high as that under high drive, when measured by the post- and pre-training log latency differences, and approximately 20 percent when measured by the number of trials to reach a performance criterion.

3. The lack of adequate conventions relating theory to data in this problem area has been noted and a justification presented for the conventions used in the present study.

4. The results have been interpreted as confirming the theoretical hypothesis (Hull's Corollary II) which was under test.

(Manuscript received August 10, 1949)

References

1. Allport, G. W. *Personality.* New York: Holt, 1937.
2. Anderson, E. E. The externalization of drive: III. Maze learning by non-rewarded and satiated animals. *J. genet. Psychol.*, 1941, **59**, 397–426.
3. Anderson, E. E. The externalization of drive: I. Theoretical considerations. *Psychol. Rev.*, 1941, **48**, 204–224.
4. Brogden, W. J. Non alimentary components in the food reinforcements of conditioned forelimb-flexion in food satiated dogs. *J. exp. Psychol.*, 1942, **30**, 326–335.
5. Felsinger, J. M., Gladstone, A. I., Yamaguchi, H. G., & Hull, C. L. Reaction latency (st_R) as a function of the number of reinforcements (N). *J. exp. Psychol.*, 1947, **37**, 214–228.
6. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97–105.
7. Finch, G. Hunger as a determinant of conditional and unconditional salivary response magnitude. *Amer. J. Physiol.*, 1938, **123**, 379–382.
8. Gladstone, A. I., Yamaguchi, H. G., Hull, C. L., & Felsinger, J. M. Some functional relationships of reaction potential (sE_R) and related phenomena. *J. exp. Psychol.*, 1947, **37**, 510–526.
9. Hanson, L. F. Master's thesis, Brown Univ., Providence, R. I., 1949.
10. Harlow, H. F. Social facilitation of feeding in the albino rat. *J. genet. Psychol.*, 1932, **41**, 211–221.
11. Horenstein, B. R. Performance of conditioned responses as a function of drive intensity. Doctoral thesis in the department of psychology, Brown Univ., Providence, R. I., 1949.
12. Hull, C. L. *Principles of behavior.* New York: D. Appleton Century, 1943.
13. Hunt, J. McV., & Schlosberg, H. General activity in the male white rat. *J. comp. Psychol.*, 1939, **28**, 23–38.
14. Hunt, J. McV., & Schlosberg, H. The influence of illumination upon general activity in normal, blinded and castrated male white rats. *J. comp. Psychol.*, 1939, **28**, 285–298.
15. Keller, F. S. Light-aversion in the white rat. *Psychol. Record*, 1941, **4**, 235–250.
16. Kimble, G. A. Personal communication.
17. Koch, S., & Daniel, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, **35**, 167–185.
18. Mote, F. A., Jr., & Finger, F. W. Exploratory drive and secondary reinforcement in the acquisition and extinction of a simple running response. *J. exp. Psychol.*, 1942, **31**, 57–68.
19. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 95–109.
20. Richter, C. P. A behavioristic study of the activity of the rat. *Comp. Psychol. Monogr.*, 1922–3, **1**, no. 2.
21. Saltzman, I., & Koch, S. The effect of low intensities of hunger on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1948, **38**, 347–370.
22. Schoenfeld, W. N., Antonitis, J. J., & Bersh, P. J. Unconditioned response rate of the white rat in a bar-pressing apparatus. *J. comp. physiol. Psychol.*, 1950, **43**, 41–48.
23. Siegel, P. S. Drive shift, a conceptual and experimental analysis. *J. comp. Psychol.*, 1943, **35**, 139–148.
24. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton Century, 1938.
25. Warner, L. H. A study of hunger behavior in the white rat by means of the ob-

struction method. Comparison of sex and hunger behavior. *J. comp. Psychol.*, 1928, **8**, 273–299.

26. Webb, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.

27. Woodworth, R. S. *Dynamic psychology*. New York: Columbia Univ. Press, 1918.

28. Young, P. T. Appetite, palatability and feeding habit: a critical review. *Psychol. Bull.*, 1948, **45**, 289–320.

29. Zener, K. E., & McCurdy, H. G. Analysis of motivational factors in conditioned behavior: I. The differential effect of changes in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, **8**, 321–350.

THE EFFECT OF REMOTE ANCHORING POINTS UPON THE JUDGMENT OF LIFTED WEIGHTS *

BY ROY K. HEINTZ

University of Maryland

Introduction

One of the many problems which has arisen concerning the functioning of frames of reference is that of determining the exact nature of the effects to be observed upon the introduction of remote anchoring points, i.e., stimuli serving to delimit the basis for judgment, remote from those being judged in terms of units pertinent to the continuum involved.

The principal impetus for studying this question derives from two experiments reported by Rogers (4). Using both inclined lines and lifted weights, he presented anchors of greater inclination and of greater weight, following a procedure previously described by Volkmann (7), and found that *O*s tended to judge the stimuli of his standard or dependently-judged series as less inclined, or as lighter, evidencing that the *O*s' absolute or subjective scales shift toward the anchoring stimuli. Results led him to conclude that "up to a certain point, the anchoring effect as measured by the shift in category thresholds is a rectilinear function of the remoteness of the anchoring stimulus. At this point there is a break in the functions and a diversity of trends appears" (4, p. 40).

Additional interest stems from McGarvey's results with verbal stimuli and verbal anchors. Reporting that results for two *O*s showed no evidence of discontinuity, she concluded that "in the results of six subjects there was *some evidence* of discontinuity" (2, p. 78; emphasis supplied).

Hunt and Volkmann (1) have shown that anchoring effects can be obtained with affective stimuli. No evidence is afforded concerning alterations in scales as a function of the distances of anchors. More recently Postman and Miller have demonstrated that anchoring effects can be observed in temporal judgments. They do not mention the matter of discontinuity, but report that "the flexibility of the scale (i.e., *O*'s absolute or subjective scale) seems to be limited. . . . The introduction of more and more remote anchors yields only diminishing returns" (3, p. 49).

Procedure

To secure further information concerning effects of distant anchors a study was undertaken using the general procedure outlined by Rogers (4), with certain differences in detail as noted below. A series of weights was presented in keeping with the Method of Absolute Judgment (or Single Stimuli); after training, progressively heavier weights were introduced as anchoring points in the latter part of each session, in regular ascending order, one only being used in each experimental session.

The standard stimuli selected for presentation throughout the experiment weighed 84, 88, 92, 96, and 100 gm. Anchors were likewise separated by equal objective differences of four gm. They weighed 100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140, 144, 148, 152, 156, 160, and 164 gm. The use of this small interval represents a material departure from the conditions of Rogers' study, intended to form a more stringent test of the hypothesis and to furnish detailed information concerning anchoring effects throughout a long range.

Encased in standard fibre containers 2.5 cm. high × 6.5 cm. in diameter, stimuli were presented as rapidly as possible by means of a revolving table, *O*'s vision being excluded by a large screen through which his wrist projected. Rest periods were allowed when requested. Instructions were similar to those given by Rogers:

> The stimuli you are to judge will be a series of different weights. You are to report their heaviness relative to each other, calling the

* This article is based upon part of a dissertation submitted to the Faculty of Princeton University in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his appreciation to Professor Hadley Cantril and Professor E. G. Wever for their guidance and direction.

lightest weights 'ONE,' 'TWO,' and so on, up to 'FIVE' for the heaviest. Report your judgment on each weight, even though the first few judgments may be difficult for you to make.

Three graduate students served as *O*s. Without prior experience and naive as to objectives, each was given several sessions of practice in which anchors were not presented. F and W had three preliminary sittings each; A had five, his introduction being interrupted. Data drawn upon the first 250 judgments in each of these sessions are reported with that for the control series; the present study also differing from that of Rogers in the respect that a control series was run throughout the investigation to secure evidence of effects carried over from one session to another, and to enable measurement of the effects of anchoring within each day's data.

To this end, subsequent sessions were divided into two portions. In the first period, constituting the *control series*, only stimuli of the standard series were presented, each being offered 50 times, in random order, for a total of 250 judgments. Immediately following, in the second portion of each sitting, anchors were introduced for the *experimental series*.

Two procedures were employed in the experimental series. Under one, anchoring points were not called to *O*'s attention, but where introduced randomly into the group of stimuli being presented, no interval of time or other cue distinguishing the series. F and W both participated under this condition, making 300 judgments in each experimental series, 50 in response to the anchor and 50 on each of the five standard stimuli, for a total of 550 judgments in each sitting.

Under the second regimen, *O* was stopped momentarily following the control period, and anchors were identified for him during the experimental period by pairing them with the standard stimuli, the latter being randomized. This was the usage employed by Rogers. The special instruction given when *O* was arrested was similar to that given his *O*s:

> The stimuli will now be presented to you in pairs instead of singly. The first member of each pair will tell you what you are to call a 'FIVE.' The second member of each pair you are to judge as before, on a scale of five, using the lower categories for the lighter weights.

A served under this condition. His task thus consisted of 750 responses per sitting: 250 during the control period and 500 during the experimental period, 250 of the latter being given in response to the anchor.

The first procedure was instituted to test the necessity for pairing stimuli with anchors, the original work by Wever and Zener (8) on the Method of Absolute Judgment suggesting that anchoring effects might be obtained without pairing.

The first anchor presented to A was 100 gm., a weight identical with the greatest stimulus of the standard series, following Rogers' usage. The first presented to F and W weighed 104 gm., representing the next unit above the greatest of the five standard stimuli with which it was randomized.

Statistical Treatment of Data

Two approaches can be employed in treating data obtained in this type of experimentation. One may calculate (a) mean values of the judgments assigned to each stimulus, or other functions of judgments assigned to each stimulus; or one may calculate (b) mean values of the stimuli to which each judgment is related, or other stimulus functions of the several judgments, e.g., category thresholds. The former approach yields functions which recede from the anchors as more distant points are presented (declining where anchors are greater than standard stimuli). Compared with mounting anchors the standard stimuli seem lighter, manifesting *contrast* (5, p. 311, *passim*). The latter approach yields functions which trend toward the anchors (rising where anchors are greater than standard stimuli). With increasing anchors, judgments at first related to lighter or lesser stimuli are accorded to stimuli in the middle or even the upper reaches of the standard series. This is a direct evidence that *O*'s scale has undergone *extension* (1, 2, 3, 4, 5, 7), and an evidence of *assimilation* (5, p. 312, *passim*).

Both modes of analysis have been employed in previous studies: Hunt and Volkmann (1) and Postman and Miller (3) using the former, Volkmann (7) the latter. McGarvey (2) and Rogers (4) present their data in both forms, Rogers placing principal reliance upon the use of category thresholds.

While the two approaches stand in logical relation, it remains to clarify their usage, for the techniques are not equally reliable nor equally serviceable; the difference appearing, moreover, to have a bearing upon the problem of discontinuity. The computation of judgment values related to the several stimuli of the standard series yields functions of equal reliability, inasmuch as each stimulus is presented with a frequency equal to that of every other stimulus. The computation of stimulus values corresponding to the various judgments, on the other hand, is less satisfactory. The several categories are not used with equal frequency. In particular,

the higher categories are used less and less as anchors increase, until they are finally forsaken. The present results are analyzed in terms of judgments reported for each stimulus in view of these considerations.

Results

For clarity of presentation in limited space, graphs in Fig. 1 for each *O* have been constructed to show the mean values of all judgments assigned to all stimuli of the standard series combined, for each control period and each experimental period. Mean judgments on anchors are also shown.[1] Results for all three *O*s reveal that judgments upon the standard stimuli were progressively altered upon the introduction of more and more extreme anchors, the tendency to respond in terms of lower categories becoming more pronounced as heavier and heavier anchors were submitted. Functions based upon them display downward trends extending to the end of the experiment, terminated with the introduction of the heaviest anchor.

It is observed that the obtained curves decline in an irregular pattern. There are recurrent fluctuations and readjustments within the general downward trend. There is no evidence, however, of any well-defined 'break' or 'breaks.' No set or series of of points shows so marked a departure as to distinguish them unmistakably from other shifts occurring in other portions of the curves. Variability is apparent throughout the data, appearing within sessions as well as between sessions; in the control series as well as in the experimental series; in the practice data as well as in the experimental data proper. The whole picture is that of a set of continuous curves showing an overall trend, varying from point to point in a random or chance pattern.

The standard stimuli were on the average judged lighter in the experimental series than in the control series. Means for the several standard stimuli in the experimental series typically fell below the respective means for the same stimuli in the control period of the same session. Differences between the functions at the established points describe the effects of anchoring within sessions.

Comparison of the graphs will disclose that the effects of anchoring tended to be greater under the condition of paired stimulation. Curves for A show greater and more consistent separation than do those for F and W.

Data collected in the control series demonstrate that anchoring effects were carried over from day to day: that experiences with anchors in earlier sessions affected judgments in subsequent sessions. Curves based upon judgments obtained therein drop in a manner analogous to curves based upon results taken in the experimental series. In F's words, "You've got me so that I expect light ones in the beginning, even before I start."

Under the randomized condition the adjudged values of anchors from the outset exceeded those for stimuli in the standard series. For F the mean values of judgments on anchors increased irregularly to the 16th session and diminished thereafter, never attaining the limiting value of 5.00, obtainable only through consistent judgments of 'five.' The curve for W reached 5.00 at the 12th session, then receded, ultimately to be established at this level from the 16th

[1] For detailed tables of the results order Documents 2738 and 2739 from American Documentation Institute, 1719 N Street, N. W., Washington 6, D. C., remitting $1.00 for microfilm (images one in. high on standard 35 mm. motion picture film) or $1.20 for photocopies (6 × 8 in.) readable without optical aid.

session. The function for A is of course constant.

F first indicated an awareness of an anchor at 116 gm., saying, "That felt like a *real* 'five.'" W shouted "'six'" upon one presentation of the 112-gm. anchor. At 112 gm., A recognized that, "The ones I am calling 'fives' are always heavier than the ones I am to judge on a scale from one to five."

Even though differences between means were reduced in the final sessions, *Os* clearly demonstrated ability to discriminate, variability of judgments notwithstanding. Means for the several stimuli were almost invariably in their proper rank orders. There were few reversals. Of 436 possible instances in which the value of a given mean might lie beyond the mean for an adjacent stimulus, only 15 cases were observed, counting four identities. Of a possible 654 second-order reversals, in which the value of a given parameter might lie beyond those of two adjacent parameters, only one case appeared.

Discussion of Results

The results are in fundamental agreement with previous studies (1, 2, 3, 4, 7) regarding the changes that occur. There is a reduction or 'lump-

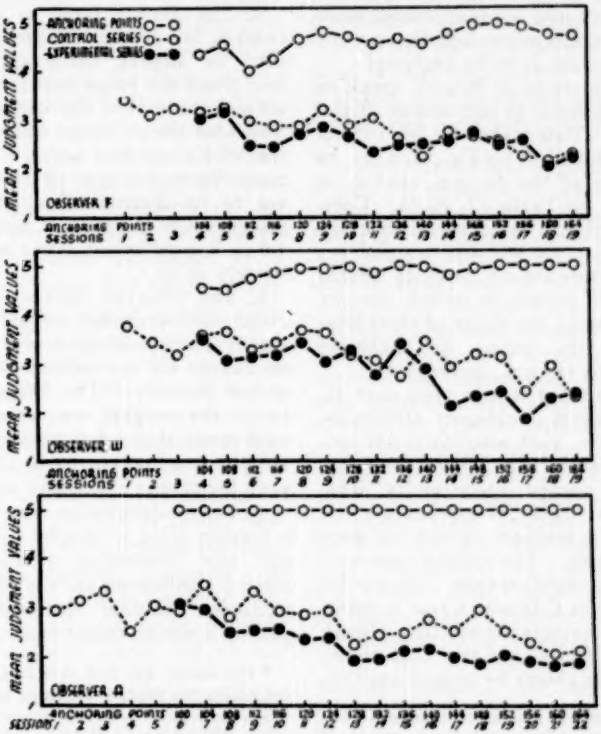


Fig. 1. Mean judged values for stimuli of the standard series as a function of the weight of anchors. Curves for the experimental series and the control series represent grand means for all standard stimuli combined. Upper curves represent means of judgments in response to anchors.

ing' of judgments in the fact of greater anchoring stimuli, signifying an extension of *O*'s scale. Changes are most marked in the case of judgments reported upon standard stimuli proximal to the anchors, as earlier studies have shown.

The outcome does not agree with Rogers' (4) findings respecting the rectilinearity, breaking, and subsequent diversity of the functions. Since details of the studies differ in several respects, the disparity might be accounted for in several ways.

1. Regression lines have not been employed in the graphic treatment of the present results. Since their use may create both an unwarranted impression of uniformity and an exaggerated sense of difference, the writer feels that a more direct treatment is to be preferred.

2. Certain errors in Rogers' graphing contribute toward an appearance of discontinuity. Two points for his subject C are not plotted in his Fig. 12, p. 33, on the ordinate of the 515-gm. anchor, as reported in his Table VI, p. 32. Comparisons with the table reveal that the regression lines for the 165, 140, and 115 gm. parameters are not properly related to their own respective points, thereby underestimating the slopes of these lines throughout the region of anchoring values 365 to 515 gm., inclusive.[2]

3. Rogers relied largely upon data expressed in terms of category thresholds. As seen above, such analysis is not uniformly reliable. The limitations are such as to *suggest* discontinuity, inasmuch as the upper thresholds are not represented in conjunction with the more distant anchors. The simpler treatment is free of this disadvantage. Graphs for Rogers' subject C furnish a case in point. His Fig. 9 convinces the eye that a break occurs by virtue of the fact that higher thresholds are absent for greater anchors. Fig. 12 seems less persuasive, even as constructed. In the former, only points for the lower two parameters might be fitted by continuous straight lines. In the latter, all six might conceivably be—especially if the errors in construction are envisaged. The writer feels it possible that the alleged discontinuity might have gone unnoticed if the more sophisticated but less serviceable mode of analysis had not been employed.[3]

4. Being more protracted, the present study affords more information concerning effects of extreme anchors. As against the use of eight anchors, curves are based upon points expressing the effects of 16 and 17 anchors, permitting one to view the shifts and fluctuations in relation to larger and more inclusive trends.

5. The anchor range is smaller *absolutely*, but greater *relatively*, than that used by Rogers, being approximately four times the range between the greatest and the least of the standard stimuli, where his anchor range was about twice that of his standard series. The anchor range reported is 15 or 16 times the interval to be discriminated, in terms of grams, where the ratio of range to interval in Rogers' investigation was approximately 10 to 1.

6. The interval between stimuli is about one fourth that utilized by Rogers. Barely discriminable, there is practically no chance for recognition of the individual stimuli. "The boundaries between the weights are vague." "It is only rarely that I feel confident about a weight; and that is only in the case of a very heavy one" (i.e., the anchor).

7. The randomization of anchors forms a further point of dissimilarity. In itself this innovation probably holds limited significance for the difference in outcomes. Results demonstrate that pairing is not necessary to secure anchor-

[2] By applying the method of least squares to data in his Table VI, the interested reader will find that lines of best fit thorugh these points will prove more nearly continuous with those through the 165 to 314 gm. anchor range.

[3] The reader will find that these statements concerning the relative merits of the two treatments are counter to Rogers' statement that the calculation of mean judgment values ". . . do[es] not alter the experimental findings in any way, save possibly to make the break in the functions more distinct" (4, p. 32).

ing effects, though the effect may be less marked than under the condition of more frequent stimulation.

It is possible that the discrepancy between the present outcome and McGarvey's report that records for a number of her *O*s show "some evidence of discontinuity" (2, p. 78) may be reconciled in terms of probable variability in the functions and the suggestive nature of one of the statistical treatments.

Results do not confirm Postman and Miller's finding that anchoring effects approach a maximum with anchors at a limited distance, with diminishing returns thereafter. The writer recognizes that it is logically possible that anchoring functions may ultimately reach a point where further changes will be of an insignificant order. Their data indicate a limit with anchors at objective distances of approximately one half (400/750ths) to two thirds of the range of their standard series. The obtained results show that changes continue to occur with anchors withdrawn to a distance four times the range of the standard series, suggesting that such a limit would only be reached with more remote anchors. Whether one can meaningfully compare such ratios is open to question, since the two continua and perceptual processes differ. Some importance may attach to the fact that the interval between their anchors 3 and 4 is not as great, proportionately, as the intervals between the remaining anchors. Coupled with uncontrolled variability, the existence of which is patent in their graphs, expected increments might fail to appear.

Results for F and W provide quantitative evidence that so-called anchoring points are not fixed points forming an invariant limit for *O*'s scale. The former never judged them consistently; the latter did not do so until relatively late, when he announced that he was "reserving judgments of 'five' for an extreme class."

Since anchors are not subjectively stable entities, the extension of *O*'s scale cannot be expected to 'follow' the regular objective increments. If it were to confrom to the anchor range with fidelity, all the standard stimuli should theoretically be judged in category 'one' in the final session, as the range of the standard stimuli (84 to 100 gm.) is then precisely one fifth of the range between the least stimulus (84 gm.) and the anchor (164 gm.). The fact that extension is incomplete was first reported by Volkmann (7).

A salient characteristic of the functions is their variability. A number of factors may be singled out as having special significance for the instability.

1. The interval is close; discrimination is difficult; consistency can scarcely be expected. The fact that variability in the practice data is of the same order of magnitude as that in the experimental data shows that the fluctuations are not a function of anchors alone.

2. Opposing forces are at work to influence judgments. Counteracting the effect of the anchor and the effect of prior experience with other anchors, one can identify another learning effect and a set engendered by the instructions. Tresselt (6) has shown that experience with one scale affects judgments upon a subsequent scale, extended practice on the first slowing adjustment to the second. Preliminary practice and continued contact with control stimuli thus theoretically serve in some degree to resist changes to be wrought by the anchors. An instructionally-derived set influences results in the same direction. Having been asked to judge on a scale from one to five, *O* attempts to distinguish between the stimuli as long as he is able to detect differences, thereby maintaining the functions at a level above that

which might otherwise be expected. Introspections reveal that *Os* were motivated to use all the categories. To consider implications, instructional set will artifactually prevent results from showing due or proper scalar extension in this type of experimentation. Momentary changes in the balance between these factors may act to produce irregularities.

3. *O*'s attitude introduces another source of variability. Disturbed by the intrusion of the anchor, he finds it difficult to maintain a consistent approach. W asserted, "After being given a series of light and medium weights, suddenly a heavy one appears. This throws my judgments off, almost makes judgment impossible. . . . I feel my judgments are being forced . . . by being required to judge the weights on a scale from one to five." He expressed a desire to use "a very much wider scale," but later reversed himself: "I should find it much easier to use a three-point or a four-point scale; five is a bit fine." Irregularities in his record beyond the 11th session undoubtedly reflect this conflict. F reported much the same thing after the 17th session. "In the beginning [control series] you give me light ones—'ones,' 'twos,' and 'threes.' It's easy to judge because you have a small frame of reference, just a few units. Toward the end [during the experimental series] you give me five units, and I find it more difficult to judge; my judgments tend to oscillate between high and low."

4. The anchoring point furnishes only a part of the frame of reference. This may be discerned in the foregoing remark, but it is revealed most lucidly in a comment by him at the 16th session. "I am trying to give you my judgments of 'four' in relation to the 'three.' It is impossible to judge it in relation to the 'five'; there is too much difference. I just try to remember what the 'three' feels like and judge what may be a 'four' on the basis of that."

Summary and Conclusions

To obtain further information concerning the effect of remote anchoring points upon frames of reference a psychophysical study was undertaken involving anchors at objective distances up to four times the length or range of a standard or dependently-judged series. Using the Method of Absolute Judgment, *Os* were given practice at judging five equidistant stimuli ranging from 84 to 100 gm., inclusive, on a response scale from one to five, light to heavy. Subsequent sessions were each divided into control and experimental periods. Anchors from 100 and 104 gm. to 164 gm. were introduced during the latter in regular ascending order, only one per session, being (a) paired with stimuli of the standard series under one condition, and (b) randomized with stimuli of the standard series under a second.

Mean judgment values were determined for each stimulus, this mode of analysis being preferred for reasons given. Results are in basic agreement with previous studies (1, 2, 3, 4, 7) in showing that heavier anchors have the effect of reducing the category values of judgments assigned to the standard stimuli, signifying an extension of *O*'s scale or frame. Collectively or severally, means for the experimental series tended to be less than corresponding means for the control series. With the introduction of more and more remote anchors the functions declined progressively, although irregularly, throughout the range studied.

In this respect the obtained results do not agree with Rogers' (4) conclusions concerning the rectilinearity, breaking, and subsequent diversity of the functions. The obtained curves appear to be continuous, subject to chance variation, with no set or series of points departing unmistakably from the larger downward trend. Results also demonstrate that

pairing is not essential, though anchoring effects may be greater under the condition of more frequent stimulation.

It follows that the results are also at variance with the report of McGarvey, who found "some evidence of discontinuity" (2, p. 78).

The outcome does not confirm Postman and Miller's (3) finding that anchoring effects approach a maximum with anchors at a limited distance from the judged series.

Despite the absence of anchors, functions for the control series tended to decline from session to session, demonstrating that present judgments on a given set of stimuli are influenced by prior experience with related anchors.

Quantitative evidence is provided demonstrating that so-called anchoring points do not have point character.

Prior experience with the standard series was considered as a factor opposing the influence of the anchors.

Instructional set was identified as a factor which may operate artifactually to prevent results from showing proper extension of *O*'s subjective scale or frame.

Introspective evidences reveal that anchoring points furnish but a part of the *O*'s frame of reference.

(Manuscript received
September 26, 1949)

References

1. Hunt, W. A., & Volkmann, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49**, 88–92.
2. McGarvey, H. R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol., N. Y.*, 1943, **39**, No. 281.
3. Postman, L., & Miller, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58**, 43–53.
4. Rogers, S. The anchoring of absolute judgments. *Arch. Psychol., N. Y.*, 1941, **37**, No. 261.
5. Sherif, M., & Cantril, H. The psychology of 'attitudes.' *Psychol. Rev.*, 1945, **52**, 295–319.
6. Tresselt, M. E. The influence of amount of practice upon the formation of a scale of judgment. *J. exp. Psychol.*, 1947, **37**, 251–260.
7. Volkmann, J. The anchoring of absolute scales. *Psychol. Bull.*, 1936, **33**, 742–743.
8. Wever, E. G., & Zener, K. E. Method of absolute judgment in psychophysics. *Psychol. Rev.*, 1928, **35**, 466–493.

PROACTIVE INHIBITION AS A FUNCTION OF RESPONSE SIMILARITY

BY ROSS L. MORGAN AND BENTON J. UNDERWOOD

Northwestern University

Introduction

The present study is concerned with the role of response generalization in proactive inhibition of verbal learning. By response generalization we mean that if Response B is associated with Stimulus A, other responses similar to B have a greater than chance probability of being elicited by A. It is assumed that the influence of response generalization can be inferred from performance changes which accompany manipulation of material along any one of several dimensions of similarity. In the present study such inferences will be made as a consequence of variation in the synonymity of adjectives.

In 1940 Gibson (1) published a comprehensive theory of verbal learning based on stimulus generalization, differentiation, and associated phenomena. At that time evidence for response generalization among verbal materials was negative, both Thorndike (8) and McGeoch (5) having been unable to demonstrate its existence. Therefore, Gibson did not include response generalization among her explanatory mechanisms. Since 1940, enough evidence, both empirical and logical, has accumulated to suggest inclusion of response generalization as a part of any stimulus-response theory of verbal learning.

At the empirical level, three studies argue for the importance of response generalization as an explanatory concept. Haagen (2), in retroactive inhibition experiments, has shown that response synonymity produced greater intercondition differences than corresponding degrees of stimulus synonymity. There was more retroactive inhibition in an experiment where similar responses were learned to the same stimuli than in an experiment where the same responses were learned to similar stimuli. In both experiments the amount of retroactive inhibition decreased with an increase in the degree of synonymity. Response generalization as the mediating mechanism is a straightforward inference.

Hughes (4) has given an empirical demonstration of response generalization in verbal learning. *S*s learned a list of 10 dissimilar adjectives as responses to 10 dissimilar nonsense syllables. One week after learning, the 10 stimuli were presented and *S* was forced to give a response to each. Judges scaled the errors which were made during the learning and recall along several dimensions of similarity to the correct responses. Errors which had the highest degree of scaled similarity to the correct responses occurred most frequently, thus providing a gradient of response generalization.

Finally, Osgood (6) has shown that interference in learning the second of two lists is less if the responses are similar than if dissimilar when the stimuli are identical in both lists.

These studies indicate that in accounting for interaction between lists, response similarity, hence response generalization, must be reckoned with.

A logical argument can be made for the importance of response similarity in learning the second of two lists when the stimuli of the two lists are identical. Maximum negative transfer (associative inhibition) occurs when a new response term is to be connected to an old stimulus term, while maximum positive transfer (associative facilitation) is obtained with continuous learning—when the same response term is connected with the old stimulus term. Since the stimulus relationships for these two situations are identical, the difference in transfer effects must lie in the response relationships.

Filling in the points between the extremes identified by the above two conditions should give a continuous change in transfer effects. Osgood (7) has recently argued for the importance of response similarity and one of the present authors (9) has given a rough theory of how response generalization would operate to produce the expected transfer effects as similarity is manipulated. This theory will be expanded in the discussion section.

In the present experiment stimuli are identical and response similarity is varied from high to low. Inferences will be made concerning the influence of generalization (1) in learning the second list and (2) in the retention of the second list after 20 min.

Procedure

Terminology.—The terms to be used in identifying lists and phenomena need specifying. On all conditions except the control condition, *S* learns two lists. These two lists will be called List 1 and List 2. Transfer phenomena which occur during the learning of List 2 will be identified as associative inhibition and associative facilitation. Transfer phenomena which occur during recall and relearning will be identified as proactive inhibition and proactive facilitation.

Basic conditions.—There are six conditions in the present experiment. Cond. A is a control condition on which a single list is learned and recalled. Conds. B through F consist of the learning of two lists and recalling List 2 after the rest interval of 20 min. Conds. B through F also have *increasing* degrees of similarity between the responses in the two lists. Thus, Cond. B is the traditional interference paradigm, A-B, A-C, in which responses are dissimilar and stimuli are identical, while on Cond. F the responses are highly similar and the paradigm may be symbolized as A-B, A-B_1. Gradations of similarity occur between these two extremes to make up Conds. C, D, and E.

List-1 learning was always carried until *S* anticipated correctly seven of the 12 responses. Likewise, List-2 learning was carried to the same criterion. After 20 min., *S* recalled and relearned List 2 to one perfect trial. Proactive inhibition will be determined by comparing the recall and relearning measures of the experimental conditions with those of the control condition (Cond. A). Associative inhibition or facilitation will be determined by comparing the learning of the single list on Cond. A with the learning of List 2 on the experimental conditions.

Materials.—The lists of 12 paired two-syllable adjectives were constructed by Haagen and used in his studies on retroactive inhibition mentioned in the introduction to this paper. Since his scaling procedure and results have been presented elsewhere (3), they need not be repeated here. From the scaled words he constructed sets of two lists each which were roughly equal on all characteristics except response similarity. The stimuli in any of the two lists of a set were identical, the responses varied in similarity. His experimental data, as well as ours, show that no significant differences in difficulty obtained between the lists of different sets taken individually. Therefore, if differences in learning List 2 of the various sets are obtained, as well as differences in its recall, they must be attributed to differences in the response similarity from set to set.

The differences in similarity may be illustrated with word pairs taken from the lists. It will be remembered that Cond. B has dissimilar responses, Cond. C slightly similar, and so on, to Cond. F, which has high similarity.

	First List	Second List
Cond. B	noiseless-sincere	noiseless-latent
Cond. C	gorgeous-agile	gorgeous-tripping
Cond. D	jellied-sullen	jellied-downcast
Cond. E	ingrown-angry	ingrown-enraged
Cond. F	willing-dirty	willing-unclean

Counterbalancing.—When the independent variable is a characteristic of the task itself, and when each *S* serves in all conditions, it is necessary to balance the tasks so that no bias due to practice and fatigue effects will enter. A form of systematic randomization was used which adhered to the following rules: (1) no two *S*s learned the conditions in the same order; (2) each condition preceded and followed every other condition four times, and (3) each condition occurred four and only four times at any given stage of practice. With six conditions 24 *S*s are necessary to carry out these principles.

For each condition, two lists were used. For 12 *S*s one specific list was used as List 1 and the other as List 2; for the other 12 *S*s the situation was reversed. Two lists also were used for Cond. A. This allowed additional practice to be given after the relearning of the single list, thus keeping the practice constant from day to day regardless of the condition. *S* recalled and relearned List 2 on Conds. B through F. Following this relearning, two recall trials were given on List 1 to avoid any possible 'sets' developing which would lead *S* to believe that the learning of the first list was unimportant.

TABLE I

MEASURES OF LIST-1 LEARNING

Cond.	Trials to Reach 7 out of 12		Total Correct Responses		Number Correct on Last Trial	
	M	σ_M	*M*	σ_M	*M*	σ_M
B	7.58	.57	22.63	1.32	7.75	.18
C	7.00	.54	23.46	1.77	7.75	.15
D	6.88	.51	23.17	1.23	7.79	.21
E	7.25	.54	24.46	1.63	7.50	.20
F	6.79	.56	23.50	1.93	7.83	.19

Miscellaneous.—The words were presented on a Hull-type drum at the standard rate of two sec. for the stimulus followed by two sec. for the stimulus and response together. Eight sec. elapsed between trials. Three different orders of each list were used to avoid serial learning.

One practice day was given during which *S* went through a complete proactive inhibition paradigm. General instructions were given on the practice day, these instructions being standard except that *S* was requested to read aloud the words as they appeared and to correct aloud any errors in anticipation.

The 20-min. rest interval between learning and recall was filled by having *S* work 'set' problems similar to those used by Luchins and working out various problems on a pyramid puzzle. These problems were counterbalanced with conditions so that if any possible interference occurred it would be the same for all conditions.

The 24 *S*s, 18 men and six women, were all enrolled in introductory psychology and were naive to verbal learning experiments.

RESULTS

List-1 learning.—Three measures of List-1 learning are shown in Table I. These are presented primarily to show that the lists of each set, when learned singly, did not differ appreciably in difficulty from those of any other set. None of the differences between any of the means for any measure is statistically significant.

List-2 learning.—An analysis of the learning of List 2 to the criterion of seven out of 12 correct responses gives the necessary evidence to evaluate the influence of response generalization on transfer. The major data are given in Table II. Cond. A, on which no prior list was learned, serves as the control for evaluating the direction of transfer effects. In terms of trials to reach the criterion it will be noted that under all experimental conditions the speed of learning is faster than on the control condition. Furthermore, it can be seen that the rate of learning is directly related to response similarity. The function is not smooth, Conds. C and D being reversed, but,

TABLE II

MEASURES OF LIST-2 LEARNING

Cond.	Trials to Reach 7 out of 12		Total Correct Responses		Number Correct on Last Trial	
	M	σ_M	*M*	σ_M	*M*	σ_M
A	7.50	.60	25.58	1.99	7.79	.18
B	6.79	.33	22.08	1.17	7.50	.15
C	5.92	.51	21.96	2.09	7.63	.18
D	6.33	.52	23.29	1.99	8.00	.24
E	5.42	.40	19.17	.96	7.79	.18
F	4.71	.37	18.08	1.34	7.88	.18

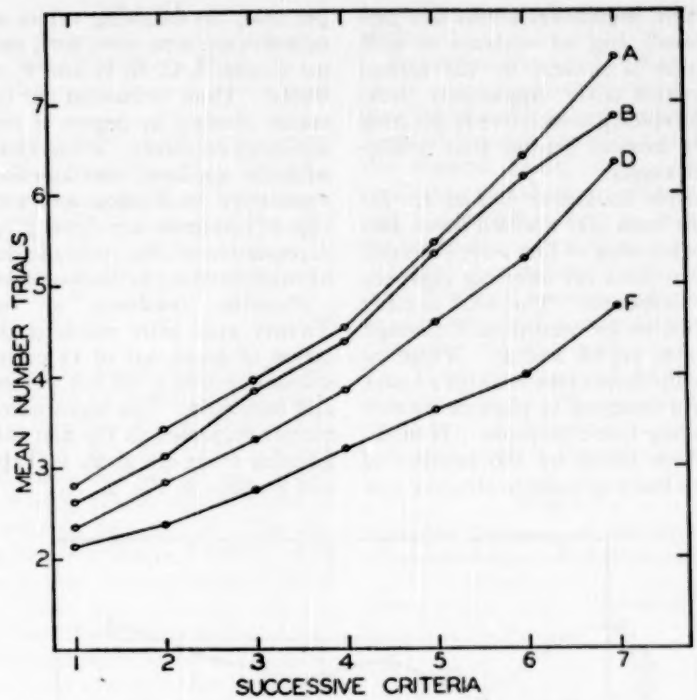


FIG. 1. List-2 learning curves showing associative facilitation as a function of response similarity. Cond. A is the control condition; Cond. B had dissimilar responses, Cond. D moderately similar, and Cond. F highly similar responses.

considering the extremes, the relationship is evident.[1] For example, the difference between Conds. B and E is $1.37 \pm .42$ trials ($t = 3.26$), which is significant far beyond the one percent level of confidence.

The overall influence of response similarity can best be seen in the learning curves for List 2. We have plotted only the curves for Conds. A, B, D, and F as samples of the similarity continuum. At no point do the curves overlap.

Cond. B is the traditional A-B, A-C paradigm. It is noted that in terms of trials to reach the criterion no associative inhibition was evident, nor does any evidence present itself for such inhibition in the learning curve. If associative inhibition is to be found, the initial trials of the second list are most likely to reveal it. A comparison of the mean correct responses on the first anticipation trial of List 1 for Cond. B with the mean correct responses at the same location of List 2 does give some evidence of associative

[1] The results throughout show reversals for Conds. C and D. We have been unable to determine the cause for this from our data, and Haagen's results (2) did not show these reversals. As a rough check on the scaling, the *S*s in the present experiment judged the words along the similarity scale used by Haagen. The mean rank orders of the scaled items were the same as found by Haagen although our *S*s tended to put the items closer together on the scale than was true of Haagen's judges.

inhibition (significant at the five per-cent level), but no evidence of such inhibition is present on the second anticipation trial. Apparently, processes producing associative facilitation quickly become greater than inhibitory processes.

Inter-list intrusions in List 2.—Intrusions from List 1 which occur during the learning of List 2 are the most pertinent data for inferring competition of responses. The total number of intrusions for conditions B through F is 7, 16, 32, 28, and 27. These include only those cases in which a List-1 response occurred in place of its corresponding List-2 response. If we divide these totals by the number of trials to learn in order to obtain a rate per trial, the following values are obtained: 1.03, 2.70, 5.06, 5.17, and 5.73, for Conds. B, C, D, E, and F, respectively. Thus, intrusions per trial increase directly as degree of response similarity increases. We are thus faced with the apparent contradiction that associative facilitation and competition of responses are directly related. A resolution of this contradiction will be considered in the discussion section.

Proactive inhibition at recall.—Twenty min. after reaching the criterion of seven out of 12 correct responses on List 2, the list was recalled and relearned. The mean number of correct responses on the first three relearning trials are given in Table III and graphed in Fig. 2.

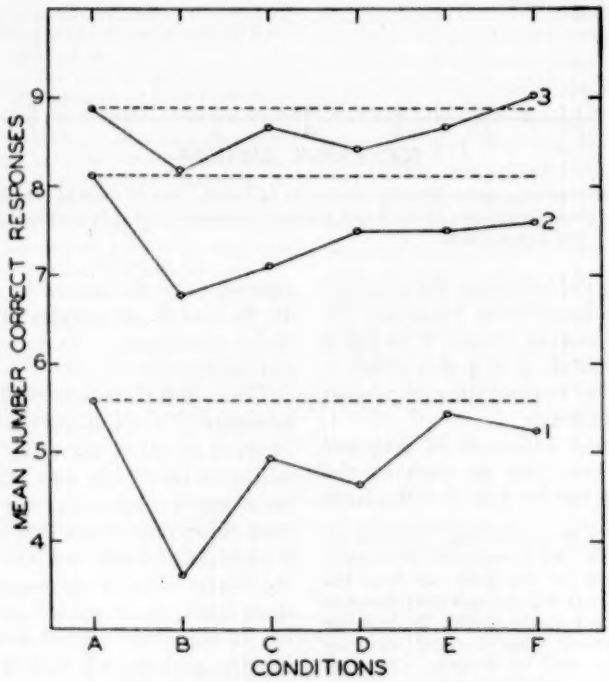


FIG. 2. Proactive inhibition as a function of response similarity. Cond. A is the control condition. Conds. B through F represent increasing degrees of response similarity. All points below the dotted lines for each of the first three relearning trials represent proactive inhibition.

TABLE III

MEAN CORRECT RESPONSES ON FIRST THREE RELEARNING TRIALS

Cond.	1		2		3	
	M	σ_M	*M*	σ_M	*M*	σ_M
A	5.58	.45	8.13	.42	8.88	.43
B	3.62	.40	6.79	.32	8.08	.36
C	4.96	.46	7.13	.32	8.67	.35
D	4.67	.35	7.50	.36	8.42	.41
E	5.46	.39	7.50	.38	8.67	.43
F	5.25	.41	7.58	.37	9.04	.31

It should be noted first that the recall scores for all experimental conditions are less than for the control condition, Cond. A. All conditions thus suggest proactive inhibition. However, reliable proactive inhibition is produced only by Cond. B. The difference between Conds. A and B on the first recall trial, $1.96 \pm .46$, gives a t of 4.26. Conds. C, D, E, and F do not differ significantly from Cond. A but all differ significantly from Cond. B. We may thus conclude that the conditions with similar responses produce significantly less proactive inhibition than the condition with dissimilar responses. The trend for the similar conditions is that the greater the similarity the better the recall. It is apparent, however, that differences at recall as a function of response similarity are much less than during the original learning of List 2. Finally, we should note that while associative facilitation occurred during learning of List 2, no proactive facilitation appeared during the first three relearning trials.

TABLE IV

RELEARNING SCORES

Cond.	*M*	σ_M
A	7.42	.71
B	7.88	.59
C	7.63	.68
D	8.12	.71
E	8.21	.80
F	7.71	.64

Relearning scores.—The mean number of trials required to relearn List 2 for all conditions is shown in Table IV. None of the differences is significant. We observe again, then, that the facilitation which occurred in learning List 2 does not appear in relearning the list to one perfect trial.

Inter-list intrusions during relearning.—The total frequency of responses from List 1 which occurred during the relearning of List 2 is shown in Table V. There are no consistent trends in the intrusions as a function of response similarity. It is noted that the condition with highest response similarity (Cond. F) produced the greatest number of intrusions. However, no appreciable difference exists among the other conditions.

TABLE V

INTER-LIST INTRUSIONS DURING RELEARNING

Cond.	Trial			Remainder	Total
	1	2	3		
B	14	2	1	2	19
C	12	3	1	2	18
D	11	2	1	5	19
E	7	2	1	5	15
F	19	12	2	5	38

DISCUSSION

This section will be concerned with an elaboration of a theory of response generalization to account partially for the obtained results.

Associative facilitation.—The first empirical generalization to be discussed is that the greater the response similarity (with stimuli identical), the greater the associative facilitation. We will first present the general theoretical approach and then bring in additional data which support it.

Assume first a gradient of response generalization which results from the learning of each item in List 1. That is, the gradient is built out from Response B as B is reinforced as a response to Stimulus A. This gradient implies that other responses, similar to B, have a greater than chance probability of being elicited by A, this probability being greater the more similar the response to B. Such a situation is depicted in Fig. 3. B_1 represents the response most similar to B, comparable to List 2 responses of Cond. F in the present experiment. B_4 represents low similarity, comparable to Cond. C., while K represents a dissimilar response in List 2, as was the case in Cond. B.

On the basis of an assumption of a gradient of response generalization, we would expect that associative facilitation would be greater the greater the similarity of the responses in the two lists. *Before S* starts to learn List 2 there is a supra-threshold tendency for Stimulus A to evoke Responses B_1, B_2, etc., this tendency being less the less the height of the gradient. In effect, because of the pre-learning of List-2 responses during the learning of List 1, less new learning will be required to reach the criterion of seven out of 12 correct responses. This pre-learning, occurring as a consequence of generalization, we shall refer to hereafter as *parasitic reinforcement.*

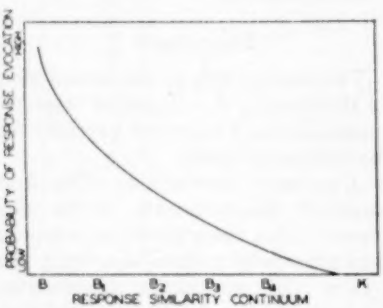


FIG. 3. Hypothetical gradient of response generalization. See text for complete explanation.

Response generalization between lists is not the only mechanism operating to produce associative facilitation in the present conditions. Since the stimuli were identical in both lists, it would be expected (according to Gibson, 1) that stimulus differentation developed in List 1 would transfer to List 2. Since the stimuli were identical in both lists of all conditions, transfer of stimulus differentation would not account for the varying amounts of facilitation observed.

It is also possible that response differentiation may take place during the learning of List 1 and that this too would transfer to List 2, resulting in facilitation. If so, it would be expected that the facilitation would be related directly to response similarity. That is, if Response B is differentiated from other responses in List 1, this differentiation should be more beneficial for learning response B_1 in List 2 than in learning Response B_4. This can be tested indirectly. If there are differences in response differentiation of List 2 for the various conditions, it should be evident in the number of *intra*-list errors. Our count of these errors shows that there are no appreciable differences in frequency of such errors among the various conditions. Thus, if transfer of response differentiation is an important factor in the facilitation observed in learning List 2, it does not show itself in the errors. Since response similarity *within* the lists was low, there is no reason to believe that the transfer of response differentiation is an important factor in the present conditions.

Additional evidence of parasitic reinforcement.—Certain empirical phenom-

ena can be deduced from our interpretation of response generalization. Tests of three such phenomena can be made from the data.

If List-2 responses are being parasitically attached to Stimulus A during the learning of List 1, we should expect that List-2 responses will be anticipated correctly sooner when (1) response similarity is high than when low, and (2) when number of reinforcements (correct anticipations) of an item in List 1 is high than when the number is low. Table VI shows that this expectation is confirmed. We counted the number of times each item was given correctly (reinforced) during the learning of List 1, and then determined 'what happened' to the corresponding List-2 item on the *first* anticipation trial. We have combined conditions in Table VI so that we have three categories of similarity (B, C-D, E-F) and three categories of List-1 response strength (0, 1–2, 3 or more). It can be seen that (1) the greater the similarity of responses the greater the percentage of correct anticipations on the first trial of List 2, and (2) the greater the number of reinforcements during List 1 the greater the percentage of correct anticipations on the first trial of List 2.

Another phenomenon which would follow from the concept of parasitic reinforcement is that a single correct anticipation on List 2 would *appear* to result in greater response strength the greater the response similarity. This would *appear* to be the case only because the higher the similarity the greater the strength of the response tendencies of the second list *before S* has had an opportunity to give the correct response. To test this we observed 'what happened' to items on the trial immediately following the trial on which they were first anticipated correctly in learning List 2. If the appearance of greater response strength per single reinforcement is to be demonstrated, the percentage repetition of response must be higher the higher the similarity of responses in the two lists. At the extremes of similarity the results are clearly in line with the proposition. On Cond. B, with 130 items observed, the percentage correct on the succeeding trial was 63.8. For Cond. F the percentage repetition was 78.5, this value being based on 121 cases. All other conditions fell between these extremes but no clear gradation as a function of similarity was observed.

A third phenomenon which might be expected from the parasitic-reinforcement interpretation is that the higher the response similarity in the two lists the more likely it is that corresponding items in the two lists will be reinforced. List 1 was learned to seven out of 12 correct responses. It is thus possible that a maximum of five items received no reinforcement during the learning of List 1. These five items would be relatively 'useless' as far as facilitating the learning of List 2 if response generalization is the facilitating mechanism. To derive maximum associative facilitation by parasitic reinforcement the items learned in List 2 should duplicate exactly the items learned in List 1. Or conversely, the items not reinforced in List 1 should cor-

TABLE VI

PARASITIC REINFORCEMENT AS A FUNCTION OF RESPONSE SIMILARITY AND FREQUENCY OF CORRECT ANTICIPATIONS

Number Correct Anticipations on List 1	Similarity of Responses								
	B (Low Similarity)			C and D			E and F (High Similarity)		
	Cases	Correct	%	Cases	Correct	%	Cases	Correct	%
0	86	2	2.3	160	11	6.9	163	13	7.9
1–2	106	6	5.7	232	38	16.4	223	64	28.7
3 and over	96	11	11.5	184	43	23.4	190	57	30.0

respond to the items not reinforced in List 2. Obviously, this extreme case would not be expected in actual practice. We might expect, however, that Cond. F would provide more instances in which neither of the corresponding items in the two lists (A-B, A-B_1) were reinforced than would Cond. B (A-B, A-K). The data confirm this expectation. Of all items not reinforced on List 1, 27 percent of the corresponding items were not reinforced on List 2, while the comparable value for Cond. F was 54 percent. The percentage values for other conditions fall in between, being 44 for Cond. C, 41 for D, and 44 for C.

We submit the above three sets of auxiliary data as supporting evidence for interpreting the associative facilitation obtained as being a function of parasitic reinforcement produced by response generalization.

Inter-list intrusions during List-2 learning.—The second major finding of the present study was that the greater the response similarity between the two lists the greater the frequency of intrusions from List 1 which occurred during learning of List 2. We thus have three directly related measures: (1) intrusions, (2) performance (associative facilitation), and (3) response similarity. Since intrusions are commonly considered to be the best index of response competition resulting in inhibition, we are faced with the problem of explaining an apparent contradiction.

In the present conditions, where stimuli were identical and response varied in similarity, we have said that the major facilitating effect which results from response generalization is parasitic reinforcement. But, we also know that a basic condition for overt competition of responses is that two or more responses be attached to the same stimulus. Furthermore, it should be more difficult to discriminate between two similar responses than between two dissimilar responses. It must be apparent, therefore, that we have a competitive situation existing in our high response-similarity conditions. The inhibition, however, which results from this competition, is less powerful than is the facilitation produced by parasitic reinforcement. The net effect of the two processes is associative facilitation.

On the basis of such an argument it may be wondered why Cond. B did not show more associative inhibition, since there would be little parasitic reinforcement involved. Two other factors making for facilitation probably account for the findings of Cond. B. The *S*s had only one practice day; hence, practice effects should provide positive transfer. Stimulus differentiation transferred from List 1 to List 2 (as discussed above) should provide some facilitating effect. These factors would, of course, produce facilitation in learning List 2 under all conditions, and while they would produce no differences in the conditions as a function of response similarity, they contribute to the direction of the transfer effects. In short, we have identified several factors which would bring associative facilitation, and one factor which would produce associative inhibition. It is reasonable that the algebraic sum of the effects of these factors would give the empirical results obtained.

Proactive inhibition at recall.—The results showed that significant proactive inhibition occurred on Cond. B. None of the other conditions produced significant proactive inhibition, but on the other hand, there was no facilitation under these conditions as was true during learning. Relatively, the recall scores showed the same relationship with response similarity as was found during learning but the differences among conditions were much attenuated. At recall there

was no consistent relationship between intrusions and response similarity. No appreciable differences in frequency or locus of intrusions was apparent for Conds. B, C, D, and E. Cond. F showed many more intrusions than did any of these other conditions.

An extension of the response-generalization theory used to account for associative facilitation would predict that proactive inhibition at recall would *increase* as a function of response similarity. The reasons for such a prediction are as follows. When an item in List 2 becomes equal in strength to its corresponding item in List 1, parasitic reinforcement should no longer facilitate the List-2 response. Since both lists were learned to seven out of 12 correct, the items would be roughly equal in response strength at the time the rest interval is introduced. According to our reasoning, therefore, parasitic reinforcement should be a minor factor at recall of List 2. Interference produced by competition of responses, on the other hand, should be greatest for the high-similarity conditions. This should lead to greater forgetting for Cond. F than for Cond. B. Since quite the opposite results were obtained, a further evaluation is in order.

What factors making for proactive *facilitation* may be present at recall in amounts directly related to similarity? If such factors can be identified, they could be said to counteract the competition factor and produce the obtained results. Two such factors are identifiable.

1. The assumption that parasitic reinforcement is of little consequence at recall is questionable. It is apparent that not all items in List 2 were equal to or stronger than corresponding List-1 responses. Therefore, some influence of parasitic reinforcement should be present at recall, and, as will be remembered, its facilitating effect should be greater the greater the response similarity.

2. A second factor reflects a methodological difficulty. This difficulty has been presented in detail elsewhere (9); hence, will be stated briefly here. The problem concerns the equality of learning of lists carried to the same mastery criterion at different speeds. It has been shown that a list learned rapidly to a given criterion is actually stronger than a list learned to the same criterion at a slow rate. If the two learning curves are extrapolated to a hypothetical 'next trial,' it can be seen that the list learned most rapidly would have a greater frequency of correct responses. In the present situation the lists with highly similar responses were learned most rapidly, would therefore be somewhat stronger than the low-similarity lists, and would give results in the direction found at recall.

Whether the facilitating influence of these two factors is great enough to counteract the inhibitory influence of competition brought on by response generalization cannot be determined from the present experiment. The recall intrusion data, which did not show consistent relationships with response similarity, argue that the situation at recall is a complex one.[3]

Relearning.—The relearning data, which showed no appreciable differences among conditions, are fairly straightforward. Since all List-2 responses should become stronger than their corresponding items in List 1 after a very few relearning trials, the influence of parasitic reinforcement of items in List 2 would be near zero. Once the List-2 responses become stronger than List-1 responses, com-

[3] Factors which we have considered as complicating the recall are as follows: (1) degree of differentiation at end of List-2 learning as a function of similarity; (2) recovery rates of generalization tendencies during rest period as a function of similarity; (3) extinction of List-1 responses as a function of similarity. These relationships (if pertinent) are not known and cannot be determined from our data.

petition will be minimized and relearning should occur at the same rate for all conditions. Actually, since recall was higher for Cond. F than for Cond. B (as illustrations of the extremes), Cond. F was relearned at a somewhat slower rate. This was probably due to the greater competition among the similar as compared to dissimilar responses.

One final analysis bearing on parasitic reinforcement will be mentioned. During the *relearning* of List 2, *List-1* responses should be strengthened by parasitic reinforcement. The strengthening as a result of such reinforcement would, of course, be less than that of List-2 responses as a result of direct reinforcement, but should be measurable nevertheless. If such parasitic reinforcement was taking place during relearning, the recall of *List 1 after* relearning of List 2 should be greater the greater the similarity of responses. In the present experiments *S* was given two recall trials on List 1 after relearning List 2 so we have an empirical check on the proposition. On the recall the mean number of correct responses for Conds. B through F was 1.7, 3.2, 3.0, 4.3, and 6.4, respectively.

Do these data support the idea of parasitic reinforcement during relearning or do they indicate merely that there was less and less interference as similarity became greater? Cond. A, the control condition, gave a mean recall of 5.6 items after 20 min. On all other conditions recall of List 1 was given after 30 to 35 min. (depending on individual learning rates). Under Cond. F the recall of List 1 was 6.4 items—appreciably higher than on the control condition although the retention interval was less on the control condition. This would argue that the gradient of recall of List 1 was not a function of less interference alone, but rather, that parasitic reinforcement of List 1 took place as a result of response generalization during the relearning of List 2.

All things considered, response generalization seems to be a very useful explanatory concept. For associative facilitation as a function of similarity it has provided a fairly complete explanation of the empirical findings; for proactive inhibition of recall it has been less complete.

Summary

The experiment was designed to determine the influence of response similarity on proactive inhibition and associative facilitation. Twenty-four *S*s learned lists of 12 paired two-syllable adjectives by the anticipation method. The degree of synonymity between response items in the two lists for each of five experimental conditions was varied systematically. On all conditions the two lists used had identical stimuli. The two lists were learned successively to a criterion of seven correct anticipations on a single trial, and the second list was recalled and relearned to a criterion of one perfect trial after a 20-min. rest interval. A sixth condition, on which a single list was learned and relearned after 20 min., served as the control condition. The results showed:

1. Associative facilitation in learning the second list increased as response synonymity increased. Increase in response similarity was also accompanied by an increase in number of intrusions from the first list during the learning of the second.

2. Proactive inhibition of recall tended to decrease as response similarity increased, although singificant proactive inhibition resulted only when responses were dissimilar. Inter-list intrusions during recall and relearning

showed no systematic relationship with response similarity.

A fairly complete interpretation of the results on associative facilitation has been given by using the concept of parasitic reinforcement brought about by response generalization. The interpretation has been less complete for the recall results.

(Manuscript received
September 15, 1949)

References

1. Gibson, Eleanor J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
2. Haagen, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Ph.D. Thesis, Univ. Iowa, 1943.
3. Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.
4. Hughes, R. A. A direct measure of response generalization. M.A. Thesis, Northwestern Univ., 1949.
5. McGeoch, J. A., & McGeoch, Grace O. Studies in retroactive inhibition: X. The influence of similarity of meaning between lists of paired-associates. *J. exp. Psychol.*, 1937, **21**, 320–329.
6. Osgood, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 227–301.
7. Osgood, C. E. The similarity paradox in human learning. *Psychol. Rev.*, 1949, **56**, 132–143.
8. Thorndike, E. L. A note on assimilation and interference. *Amer. J. Psychol.*, 1937, **49**, 671–676.
9. Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, Inc., 1949.

REACTIVE INHIBITION AS A FUNCTION OF NUMBER OF RESPONSE EVOCATIONS

BY PAUL S. SIEGEL

University of Alabama

It is a common observation, both formal and informal, that organisms often 'avoid' the repetition of a reaction. To cite a few representative studies, Thorndike (19) has reported that human subjects usually fail to repeat a voluntary response upon the repetition of the original stimulus. Similar observations have been made by Luh (11), by Telford (17) and by Telford and Thompson (18). Further, animal studies have shown that the failure to repeat often occurs when the reaction is highly adaptive, i.e., fulfills known needs. The experiments of Dashiell (2), of Dennis (3, 4), of Dennis, Henneman, and Varner (5), Dennis and Sollenberger (6), and of Wingfield and Dennis (20) are illustrative. Pavlov (14) and Hovland (9) have both demonstrated a paradoxical decrement in the strength of a conditioned response following its reinforcement and Hovland has termed this the 'inhibition of reinforcement.'

Recently, Miller (12) and Mowrer (13) have proposed a general explanatory principle which has been termed the Mowrer-Miller hypothesis. This principle, which states that the occurrence of a reaction may weaken or inhibit the tendency to make the reaction (somewhat like the action of fatigue), has found its most rigorous formulation within the system of adaptive behavior developed by Hull (10). Here the principle is identified as the postulate of reactive inhibition (Postulate 8) and is expressed as follows:

> Whenever a reaction (R) is evoked in an organism there is created as a result a primary negative drive (D); (a) this has an innate capacity (I_R) to inhibit the reaction potentiality (${}_SE_R$) to that response; (b) the amount of net inhibition ($\dot{I}_R$) generated by a sequence of reaction evocations is a simple linear increasing function of the number of evocations (n); and (c) it is a positively accelerated function of the work (W) involved in the execution of the response; (d) reactive inhibition (I_R) spontaneously dissipates as a simple negative growth function of time (t''') (10, p. 300).

This postulate is employed by Hull in the derivation of such behaviorally diverse phenomena as disinhibition, spontaneous recovery, reminiscence, the economy of distributed practice, and the law of less work or effort. Part (c) of the principle finds substantiation in the experimental studies of Mowrer and Jones (13) and Crutchfield (1). Part (d) is in part supported by a study performed by Heathers (7). This paper proposes an examination of Part (b).

It is assumed that when two well established but physically incompatible reactions (R_1 and R_2) are simultaneously evoked in a series of stimulus presentations, frequency of occurrence (p) may be regarded as an index of the strength of the reaction tendency (R_1 and R_2).[1] When this is taken in conjunction with Part (b) as stated above, it may be reasoned that, other conditions constant, a decrease in the frequency of occurrence of R_1 should follow its evocation. This comes about through a weakening of R_1 resulting from the action of reactive inhibition. R_2 enjoys a momentarily greater relative strength and will occur with greater frequency. Further, the weakening of R_1, reflected in the increased frequency of occurrence of R_2, should be directly proportional to the number of prior evo-

[1] This is an amalgamation of several of Hull's postulates (10): notably, 11, 12, and 16.

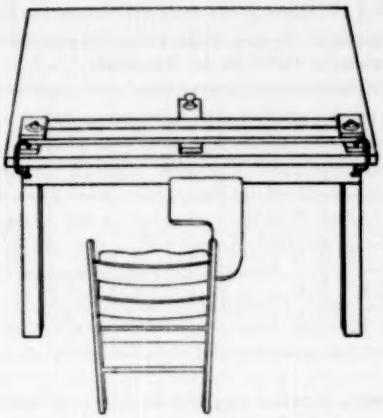


FIG. 1. Front view of apparatus

cations of R_1. The increase in the frequency of occurrence of R_2 is here conceived to be a direct behavioral reflection of the growth of reactive inhibition.

METHOD

Subjects and apparatus.—Volunteers (250) were recruited from the undergraduate population of the University of Alabama.[2] None had completed advanced work in psychology, the sex distribution was 88 percent male and 12 percent female, ages ranged from 17 to 40 yr., and all *S*s were predominantly right-handed.

The apparatus (see Fig. 1) consisted of two ordinary light switches selected for similarity of appearance, and spring tension, mounted on a table 30 in. apart with a five-watt bulb stationed midway between and slightly beyond them. The switches were wired in series with the bulb and a remote key operated by *E*. The *S* was seated at a desk-arm type chair within convenient reach of both switches.

Preliminary preparations consisted of having the seated *S* laterally adjust the chair until satisfied that the two switches were equally easy to reach with the right hand. The general instructions were then read by the *E*:

> Before you is a light bulb and two switches. When the light goes on, please turn it off, using either switch. They are identical and both will turn off the light. It does not matter which one you use. After turning off the light, return the switch to its normal position and bring your hand back to its original position. I can answer no questions about the nature of the experiment. Are the instructions clear?

Each *S* was then practiced four times in the sequence, RLLR. The following instructions were then read by *E*:

> This experiment is divided into two parts. During the first part, please use only the *right* switch. During the second part, you are free to use *either* switch. I shall tell you when the second part begins. Remember, during the second part you are free to use *either* switch.

*S*s were then given varying amounts of exercise with the right switch. At this stage, five groups ($N = 50$) must be distinguished according to the number of right responses given:[3] Group 0—0 right reactions; Group 20—20 right reactions; Group 40—40 right reactions; Group 80—80 right reactions; Group 160—160 right reactions.

During this series, the *S* set his own tempo. As soon as the reaction had been made and the *S* had returned his hand to the resting position, *E* presented the stimulus again. The average rate of presentation was approximately one each two sec. At the end of the series, a two-sec. break was interpolated during which the further instructions were read:

> You may now choose either switch.

A free-choice series of 10 stimulus presentations was immediately given at the same tempo. Following this the *S* was asked to account for his choices during the last series and dismissed.

In the instance of Group 0, the absence of exercise with the right switch necessitated a slightly altered procedure. To the general instructions was added:

> Remember, each time the light goes on, turn if off using whichever switch you prefer. Return the switch to its original position and bring your hand back to rest.

Practice was then given in the RLLR sequence, the instructions, "You may now choose either switch," were read, and the free-choice series immediately initiated.

RESULTS AND DISCUSSION

The number of left responses yielded by the subjects of each group during

[2] Mr. Morris Steinberg of this department assisted in the conduct of the experiment.

[3] The same sex distribution was observed in each group. However, an earlier and similar study (15) suggests no genuine sex difference in this situation.

TABLE I

FREQUENCY DISTRIBUTIONS OF LEFT-CHOICE RESPONSES DURING FREE-CHOICE SERIES OF 10 RESPONSES AS A FUNCTION OF NUMBER OF PRIOR RIGHT RESPONSES

No. Prior Right Responses	0	1	2	3	4	5	6	7	8	9	10	Mean	p
0	17	0	3	5	4	7	5	5	1	0	3	3.5	.35
20	5	7	1	4	8	8	5	6	3	0	3	4.4	.44
40	7	5	5	3	2	11	5	6	3	2	1	4.2	.42
80	10	1	2	1	2	7	11	6	3	2	5	5.0	.50
160	2	0	0	2	3	6	9	6	8	4	10	6.9	.69

the free-choice series is found in Table I. The table is easily read. For example, the first entry (top, left) indicates that 17 subjects of Group 0 gave no left reactions during the free-choice series of 10 stimulus presentations. The other entries may be similarly interpreted. The mean for each group has been entered in the table as well as the simple proportion of left responses (frequency of occurrence).

A consideration of the means (or the proportions) of Table I indicates a decided trend. As prior exercise in the right reaction increases, the frequency of occurrence of the left response increases; presumably, through the weakening of the right reaction through reactive inhibition. Applying the chi-square test of independence to the data, a value of 67.8 is obtained. For 24 degrees of freedom, this is significant at a level well beyond the one percent fiducial limit.[4] We may safely conclude that it is highly improbable that the number of left responses is independent of number of prior right reactions, i.e., statistically significant differences exist among the groups.

Frequency of occurrence (p) has been plotted against number of prior reactions in Fig. 2. Inspection of the empirical plot strongly suggests rectilinear regression.[5] Employing the method of least squares, the straight line function, $p = .002\ n + .36$, has been fitted to the data. The fit is seen to be a reasonable one. It is of course, impossible to prove that this relationship is a straight line. We can merely show that the obtained deviation from rectilinearity is easily attributed to random or chance fluctuation. The test for rectilinear regression described by Snedecor (16, p. 382) has been applied to the data and is summarized in Table II. The obtained F-value of .44 indicates that deviations of the magnitude here observed may be readily attributed to chance fluctuation about a true straight line.[6] Regarding the increase in p as a direct

[4] Sixteen degrees of freedom were lost in the combination of frequencies for categories containing less than five cases.

[5] Like the means, the medians also seem to fall along a straight line. Means were elected here because of the ease of calculating the statistical significance of the deviation from linearity.

[6] It is unfortunately true that this test assumes the normality of the underlying parent population. Consideration of Table I reveals that several of the distributions are quite skewed. There seems to be no normalizing transformation that would not overcorrect some of the distributions. As derived from these data, this statistic contains an undetermined degree of error. It is hoped that the use of a large number of cases has minimized this error.

reflection of the growth of reactive inhibition, Hull's principle is clearly verified. In view of the fact that a a voluntary reaction of the human subject was measured, the results are remarkably uniform and the agreement quite striking. Perhaps this uniformity will lend substance to the most recently voiced plea by Hilgard (**8**) for a more naturalistic approach in psychological experimentation.

It is interesting to inquire into the locus of reactive inhibition. The two gross possibilities that present themselves are peripheral and central. Mowrer and Miller have plainly adopted a peripheral view. Reactive inhibition is conceived to be a state of muscular fatigue or strain, or both. Hull's position is less clear, perhaps for the reason that the problem does not arise in a strictly molar system. It is to be noted that his formal postulate contains no assumption as to locus. His sub-molar considerations seem to imply acceptance of the Mowrer-Miller hypothesis or ultimate peripheral localization. The terminal reports of our subjects are interesting in this context. Although illuminating, no definitive answer emerged from an analysis of these introspective reports. Quite marked in the instance of the 80 and the 160 groups, was the frequent explanation of predominant left choice as resulting from muscular fatigue or strain.[7] Typical reports were:

"'Cause my hand was tired."
"My arm was tired from going to the right one."
"Hand tired from going right."

Clearly, a peripheral locus is indicated. However, equally frequent were such reports as:

"To break monotony."
"To vary the routine."
"Moved to left for variety."

A more central involvement seems indicated. In the last analysis, the inhibition here observed probably has both peripheral and central aspects.

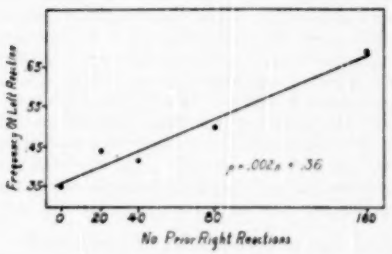


FIG. 2. The relationship between number of left choices made during the free-choice series and number of prior right reactions

TABLE II

TEST FOR LINEAR REGRESSION

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
No. Prior Right Reactions	4	329	
Linear Regression	1	318	
Deviations from Linear Regression	3	11	3.7
Error	245	2063	8.4

$$F = \frac{3.7}{8.4} = .44$$

SUMMARY

Hull's principle of reactive inhibition, which states that a reaction is

[7] Extrapolation of the curve indicates that the frequency of occurrence of the left reaction would become 1.00 following 320 prior right reactions. Beyond this point, frequency of occurrence will not reflect the further accumulation of reactive inhibition. Some other behavioral index must be employed. Although complaints were rare in our 160 group, this number of consecutive right responses approaches the endurance limit. The experimenter interested in an empirical check of the above prediction will be well advised to provide himself with means for countering physical aggression.

weakened in direct proportion to the number of times it is evoked, was examined in an experimental situation affording the human subject the opportunity for making either of two alternative responses. It was found that the frequency of occurrence of one response increases with the amount of prior exercise in the execution of the other. Agreeing with Hull's principle, this increase was found to be rectilinear.

(Manuscript received August 26, 1949)

References

1. Crutchfield, R. S. The determiners of energy expenditure in string-pulling by the rat. *J. Psychol.*, 1939, **7**, 163–178.
2. Dashiell, J. F. Direction orientation in maze running of the white rat. Comp. Psychol. Monog., 1930, **7**, No. 2, 1–72.
3. Dennis, W. A comparison of the rat's first and second explorations of a maze unit. *Amer. J. Psychol.*, 1935, **47**, 488–490.
4. Dennis, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, **28**, 305–312.
5. Dennis, W., Henneman, R. H., & Varner, W. B. A further study of the initial maze behavior of rats. *J. genet. Psychol.*, 1932, **41**, 358–368.
6. Dennis, W., & Sollenberger, R. T. Negative adaptation in the maze exploration of albino rats. *J. comp. Psychol.*, 1934, **18**, 197–206.
7. Heathers, G. L. The avoidance of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10**, 359–380.
8. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
9. Hovland, C. I. 'Inhibition of reinforcement' and phenomena of experimental extinction. *Proc. nat. Acad. Sci., Wash.*, 1936, **22**, 430–433.
10. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
11. Luh, C. W. The avoidance of repetitive responses. *Psychol. Rev.*, 1931, **38**, 276–279.
12. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
13. Mowrer, O. H., & Jones, H. M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369–386.
14. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927.
15. Siegel, P. S., & DeYampert, A. Conditions of human variability. (Unpublished)
16. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.
17. Telford, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1–36.
18. Telford, C. W., & Thompson, N. Some factors influencing voluntary and reflex eyelid responses. *J. exp. Psychol.*, 1933, **16**, 524–539.
19. Thorndike, E. L. The refractory period in associative processes. *Psychol. Rev.*, 1927, **34**, 234–236.
20. Wingfield, R. C., & Dennis, W. The dependence of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135–147.

THE INFLUENCE OF THE INTER-TRIAL INTERVAL ON THE HUMPHREYS' 'RANDOM REINFORCEMENT' EFFECT DURING THE EXTINCTION OF A VERBAL RESPONSE [1]

BY DAVID A. GRANT, JOHN P. HORNSETH, AND HAROLD W. HAKE

University of Wisconsin

The object of the present experiment was to investigate the influence of the inter-trial interval on the Humphreys' 'random reinforcement effect' during the extinction of a verbal response.

Humphreys, Miller, and Ellson (6) observed that variation of the inter-trial interval from 5 to 20 sec. had little effect upon the acquisition, extinction, or spontaneous recovery of verbal expectations. These investigators remarked that more information was needed on expectancy and decremental factors in conditioning before definite conclusions could be drawn as to the role of expectancy (or similar processes) in conditioning. The growing literature on partial or intermittent reinforcement (7, pp. 522 ff., 4, pp. 112 ff., 3, 2, 9) sustains interest in their remarks, and the present study extends their findings.

Apparatus and Procedure

The apparatus was similar to that of Humphreys (5). Two frosted 60-watt bulbs were mounted on the left and right sides of a three-foot square, vertical, black board. The lights were controlled by a concealed electronic timer. The left-hand light served as the 'conditioned stimulus' (CS), and the right-hand light served as the 'unconditioned stimulus' (UCS). A given trial was considered a 'reinforced' or *positive trial* if the right-hand light flashed on at the termination of the left-hand light. The interval between onset of the left-hand flash and the occurrence of the right-hand flash was three sec. *Negative or non-reinforced trials* also were given in which only the left-hand light was flashed.

The *Ss* were 184 men and women students who volunteered from elementary psychology classes. They were all read standard instructions directing them to guess on each trial whether the right-hand light would flash on. They were instructed to record their responses on IBM answer sheets immediately after the left-hand light came on. The *Es* checked carefully to see that the *Ss* followed these instructions. Experiments were carried on with groups of *Ss* in small lecture rooms.

A tri-factorial experimental design was used as outlined in Table I. The three factors were: (a) spaced *vs.* massed acquisition trials; (b) spaced *vs.* massed extinction trials; and (c) 50 percent random *vs.* 100 percent positive trials or 'reinforcement.' The interval between massed trials was about five sec., and the interval between spaced trials was about 40 sec. In the 50 percent 'random reinforcement' procedure half of the acquisition trials were positive and half negative. The positive trials were scheduled randomly except that no more than two consecutive positive or negative trials were permitted. Twenty-three *Ss* were used with each of the eight procedures called for by the design.

Results

The *S* could make two types of response. If he guessed that the second light would flash on, it was called a *positive response*. If he guessed that the second light would not flash on, it was called a *negative response*. The essential findings of the experiment are readily apparent in Fig. 1. In this figure are plotted the acquisition and extinction of the positive responses for the eight groups of *Ss*. The data are averaged in percent frequency in successive five-trial blocks. The figure shows clearly that all four groups receiving 100 percent positive trials

[1] This research was supported in part by a grant from the Research Committee of the Graduate School of the University of Wisconsin from funds provided by the State Legislature for 1948–49.

TABLE I

DESIGN OF THE EXPERIMENT

Procedure	Acquisition 50 Trials	Extinction 20 Trials
A	100% Massed	Massed
B	100% Massed	Spaced
C	100% Spaced	Massed
D	100% Spaced	Spaced
E	50% Random Massed	Massed
F	50% Random Massed	Spaced
G	50% Random Spaced	Massed
H	50% Random Spaced	Spaced

during acquisition show a rapid rise in their frequency of positive responses, and all four groups receiving 50 percent positive trials continue to emit about 50 percent positive responses throughout acquisition. During 'extinction' the frequency of positive responses drops rapidly in all four 100 percent 'reinforcement' groups, but the drop is much more gradual in the four 50 percent 'reinforcement' groups. The only factor which clearly affects the responses is the partial 'reinforcement.' Inter-trial interval has no consistent effect upon acquisition, extinction, or upon the partial 'reinforcement' phenomenon.

The differences between groups were tested by means of analysis of variance. Two acquisition scores were obtained for each *S* by totalling separately the positive responses during the first 25 trials and second 25 trials of acquisition. An extinction score was obtained for each *S* by totalling the number of positive responses during the extinction trials. Although these scores were not as continuous as might be desired, the results of the analyses were sufficiently striking to overcome this difficulty. The analyses are summarized in Table II. (The acquisition analyses involve only two factors since the spacing or massing of extinction is irrelevant.) With an *F* of 70.39 during trials 1–25 and 500.46 during trials 26–50, the partial 'reinforcement' effect is highly significant. Inter-trial interval produces no significant effect upon acquisition, in general, or upon the partial 'reinforcement' phenomenon as shown by the insignificant *F*s for interactions.

The analysis of the extinction scores is equally clear. Partial 'reinforcement' affects extinction, as is shown by an *F* of 158.92, but inter-trial

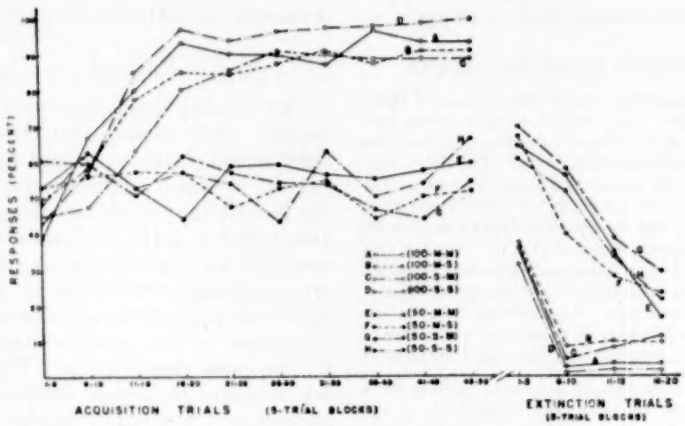


FIG. 1. Acquisition and extinction curves of the Humphreys' verbal response. The percentage of positive responses is plotted for successive five-trial blocks for each of the eight experimental groups.

TABLE II

ANALYSES OF VARIANCE OF THE ACQUISITION AND EXTINCTION SCORES

Source of Variation	Acquisition Trials 1–25	Acquisition Trials 26–50	Extinction
Partial *vs.* 100% Reinforcement	70.39*	500.46*	158.92*
Acquisition Massing	1/75.70	1/33.31	1/19.30
Extinction Massing			1/1.33
Reinforcement × Acquisition	1.33	1.16	1.01
Reinforcement × Extinction			1/2.45
Acquisition × Extinction			1/1.09
Reinforcement × Acquisition × Extinction			1.41
Replication (Error estimate)	(9.84)**	(8.66)**	(10.42)**

* Significant at the 0.1% confidence level.

** Error variance estimates. All other entries in the table are *F*s obtained by dividing the mean squares (in parentheses) into the mean squares for the sources of variation listed in the left-hand columns.

interval has no apparent effect upon extinction regardless of the spaced, massed, partial or 100 percent character of the previous acquisition trials.

DISCUSSION

Inter-trial interval had little or no effect upon acquisition or extinction of the verbal expectations of the present experiment. This corroborates and extends the results of previous investigators (**6**). The present experiment like that of Sheffield (**9**) was designed, incidentally, to separate the effects of spacing acquisition trials from the effects of spacing extinction trials, so that the present conclusions have a broader basis than did those of the earlier paper (**6**). The absence of significant interactions showed further that the inter-trial interval had relatively little effect upon these differences in acquisition and extinction which are introduced by partial 'reinforement.' These results differ markedly from those of Sheffield (**9**) who used the running response of rats in a single alley.

Humphreys, Miller, and Ellson reasoned that phenomena determined by expectancy in different experiments would show similarities. Further, they argued that the work decrement factor was of considerable import in bringing about differences between spaced and massed conditioning trials such as those reported by Calvin (**1**) and Reynolds (**8**). In view of the absence of decremental effect on the partial reinforcement phenomenon in the present study, this line of reasoning leads to the prediction that massing or spacing acquisition and extinction of a CR such as the conditioned eyelid response would have little effect on the verbal expectancy aspect of the eyelid responses under partial reinforcement. In the event that inter-trial interval is found to influence the relative amount of the Humphreys' effect in eyelid conditioning, one of three interpretations would be indicated: (a) the 'decremental' effect and the 'expectancy' effect are mutually dependent; (b) these effects are non-linear; (c) the 'decremental' and 'expectancy' effects are non-linear and mutually dependent. It seems clear that the influence of inter-trial interval phenomena will be found to vary from one action system to another and perhaps from species to species (**9**) so that general theoretical interpreta-

tions of these phenomena are premature.

Summary

Verbal expectations of 184 college students were studied using Humphreys' technique in a tri-factorial experimental design in which (1) inter-trial interval during acquisition, (2) inter-trial interval during extinction, and (3) 100 percent *vs.* 50 percent 'partial reinforcement' were contrasted. Fifty acquisition trials and 20 extinction trials were used. The principal results were as follows:

1. The work of Humphreys, Miller, and Ellson was corroborated and extended. Inter-trial interval had no effect upon acquisition or extinction of the verbal expectations.
2. A clear-cut partial 'reinforcement' or Humphreys' effect was obtained.
3. Inter-trial interval had no effect upon the Humphreys' partial 'reinforcement' phenomenon.

The results suggest an interpretation of the Humphreys' phenomenon largely in terms of non-decremental or verbal factors as opposed to decremental factors. It is suggested that if inter-trial intervals affect the relative degree of the partial reinforcement phenomenon with the eyelid CR, or other action systems which are subject to decremental factors, the decremental and non-decremental factors interact or are non-linear. That a general theoretical treatment of partial reinforcement phenomena is premature is illustrated by the fact that the present findings differ markedly from those of Sheffield on a different response in a different species.

(Manuscript received
September 24, 1949)

References

1. Calvin, J. S. Decremental factors in conditioned response learning. Ph.D. dissertation, Yale Univ., 1939, cited in: Hilgard, E. R., & Marquis, D. G., *Conditioning and learning*, New York: Appleton-Century, 1940, p. 362.
2. Grant, D. A., Meyer, D. R., & Hake, H. W. Proportional reinforcement and extinction of the conditioned GSR. *J. gen. Psychol.*, 1950, **42**, 97–101.
3. Grant, D. A., Riopelle, A. J., & Hake, H. W. Resistance to extinction and the pattern of reinforcement: I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.
4. Hilgard, E. R. *Theories of learning*. New York: Appleton-Century, Crofts, 1948.
5. Humphreys, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
6. Humphreys, L. G., Miller, J., & Ellson, D. G. The effect of the inter-trial interval on the acquisition, extinction, and recovery of verbal expectation. *J. exp. Psychol.*, 1940, **27**, 195–202.
7. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
8. Reynolds, B. R. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
9. Sheffield, Virginia F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

THE RELATIONSHIP BETWEEN LENGTH OF INTERVAL SEPARATING TWO LEARNING TASKS AND PERFORMANCE ON THE SECOND TASK[1]

BY CHARLES E. HAMILTON[2]

From the Psychological Laboratory, State University of Iowa

The background of experimental work and thought regarding the variable of warm-up effect in human learning has been presented by Thune (5) in a preceding, collateral study and need not be reviewed in detail here. In the verbal learning situation, by analogy with the motor learning situation as analyzed by Ammons (1), it may be presumed that a state or condition of warm-up is acquired by subjects during practice on the learning task. When rest follows, the acquired warm-up state may be partially or completely lost during this time and must be regained during the course of subsequent practice on the same or similar tasks. The state of warm-up itself is thought to consist of the appropriate attitudinal and postural adjustments adopted by subjects to accomplish the task. Such adjustments are acquired over a number of trials rather than instantly and are lost gradually, rather than suddenly during the rest period following. Thus, after practice on one task, subsequent performance on the same or a similar task should benefit as a direct function of the amount of practice given the first task immediately prior to the second, and inversely as the amount of time separating the two tasks is increased. These views indicate two separate but closely related experimental problems demonstrating their validity. The first, that of showing an increased level of performance on a second verbal learning task following different amounts of immediately prior practice on the first (rest interval constant) has been presented by Thune in a preceding study. The second, that of showing a decreased level of performance on a second verbal learning task following different amounts of rest separating this task from the first forms the subject for the present study.

Irion (3, 4) studying retention as a function of the length of rest interval presents evidence which indicates that part, at least, of the loss in retention observed with increased lengths of rest may be attributed to loss of warm-up adjustments rather than loss in the form of 'forgetting.' He tested recall of subjects for a list of paired-adjectives learned for ten trials and relearned after intervals of 0 (massed), 1, 5, 10, 20, or 40 min., or 24 hours. He found that recall, as measured by the first relearning trial, decreased as the length of the rest period increased. To test whether such loss, seen in similar situations since the time of Ebbinghaus, could be attributed to loss of warm-up 'adjustments,' he introduced a condition intended to reinstate some degree of the postulated warm-up without formal learning taking place. This condition was intended for comparison with the other conditions to determine whether as great loss in retention would be observed. This was accomplished by having one group of subjects duplicate the 24-hour condition with the exception that immediately prior to the first relearning trial (after 24 hours rest) they were given one trial of naming colors exposed at the window of the memory drum in the same manner and rate as the words themselves, then immediately were given the relearning trials. Recall scores under this condition were superior to similar scores for each condition of rest from five min. to 24 hours and very nearly equal to the one-min. rest condition. From this Irion concluded that the observed loss in re-

[1] Based on a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The author is greatly indebted to Professor Kenneth W. Spence who directed the investigation.

[2] Now at Montana State University

tention following rest under the various conditions was due at least in part to the loss of warm-up adjustments acquired during learning and the beneficial effects of the color-naming trial resulted from the reinstatement of some of these adjustments.

Irion's results suggest that it is possible to demonstrate a parallel sequence of events in the situation in which new verbal material is learned following the learning of a first, similar material, with a variable period of rest introduced between the two learning tasks. Data relevant to this problem may be found in the study by Thune already cited. Among his experimental conditions is one which involves an interval of 24 hours between practice on an earlier, warm-up list and the learning of the second test list. A comparison of the rate of learning of the second list under this condition with that under the condition in which an equal number of trials were given on the warm-up list only 10 minutes prior to the test list, showed that the rate of learning following the longer lapse of time was significantly inferior to that following the shorter lapse of time. Since under these conditions new material was being learned following rest (rather than previously learned material being recalled), the evidence for the operation of warm-up effect (the loss of such adjustments) is relatively uncomplicated. The demonstration by Thune of the adverse effect upon learning of a 24-hour rest as compared with a 10-min. rest indicates limits within which the more precise nature of this functional relationship may be investigated.

Purpose of Experiment

The present experiment was designed to investigate the nature of the relationship between level of performance on a verbal learning task following previous practice on a similar task and different lengths of interpolated rest interval of less than 24 hours. To this end, nine groups of *Ss* received 10 trials on a list of paired adjectives (to be referred to as the warm-up list) followed by a rest interval of 8 sec., 5, 10, 20, 40, 60, 120, 180, or 240 min., and in turn followed by 10 trials on a list of paired nouns which served as the test list. Results obtained in this fashion should provide a relatively direct measure of the extent to which warm-up adjustments acquired during practice on the first list are lost during the period of rest following, as shown by level of performance on the test list. Knowing from Thune's data that such a loss is significant when a period of 24 hours intervenes, these data should provide information as to whether such loss is uniform and gradual over this period or possesses some other functional characteristic.

Experimental Technique

Since the present study was conducted in close coordination with the preceding study by Thune, effort was made to duplicate insofar as possible experimental features of the two investigations. To this end, the learning materials employed, total number of learning trials on lists, and general procedures for the two studies are the same; specific differences lie in varying either the amount of immediately prior practice on the warm-up list, rest interval constant (Thune's study), or holding practice constant and varying the rest interval (the present investigation). For these reasons data so obtained are directly comparable from one study to the other.

Apparatus and materials.—A Hull-type memory drum was employed to present the paired-associate learning material. This material was exposed at approximately a four-sec. rate including the two-sec. anticipation interval. An eight-sec. interval was employed between successive presentations of a list. The memory drum was mounted in a screen so that *E* could be seated behind it and not be seen by the *S*. The exposure window of the drum was lighted by a shaded 7½-watt electric lamp. Lists of 15-unit paired associate one-syllable nouns and two-syllable adjectives were employed as learning materials. These were presented to *Ss* in three different orders to avoid serial position effects. The words for the lists were selected according to conventional rules for list construction; in

addition, an attempt was made to use pairs as nearly alike in associative value as possible. Materials employed during the rest interval introduced under certain of the conditions consisted of cartoons clipped from various periodicals and mounted on 5 × 7 cards.

Subjects.—A total of 225 *Ss* served in the experiment; each *S* served under a single experimental condition only, so that each of the nine experimental groups consisted of 25 *Ss*. All *Ss* were obtained from undergraduate courses in psychology; none had served previously in a memory drum experiment and all were naive with respect to the purpose of the experiment. *Ss* were discarded from the experiment if they failed to meet certain criteria for minimal and maximal performance: (1) failure to make a correct response during either preliminary practice trials or during the ten trials on the warm-up list, (2) in the case of the last five *Ss* in each group, failure to make at least 10 correct responses on the warm-up list, or (3) accomplishment of a perfect score on one of the first five trials of the warm-up list. The total number of *Ss* so discarded was 31. *Ss* were assigned to the conditions in a partially random fashion; random assignment was made within the conditions employing rest intervals of (1) from 8 sec. through 10 min., (2) 20 and 40 min., and (3) 60 min. through 240 min.

Procedure.—Two days of testing were employed. On the first day general instructions were given and *Ss* learned a practice list of paired one-syllable nouns for six trials. This served to acquaint the *Ss* with the experimental situation and reduce variability due to differences in understanding instructions, etc. The experiment proper was conducted on the following day. At this time *Ss* in conditions employing rest intervals up to and including 40 min. had one experimental session in which they first learned a list of two-syllable adjectives (the warm-up list) for 10 trials. This was followed by a period of rest according to the condition assigned; at the conclusion of the rest interval the test list consisting of a new group of one-syllable paired nouns was learned for 10 trials. During the rest interval *Ss* were required to sort cartoons described previously. *Ss* serving under conditions employing rest intervals over 40 min. were dismissed from the experimental room during the rest period. No restriction was placed on their activity during this period aside from a caution to not discuss the experiment with other individuals. All questions were deferred until the end of the experiment. Scores were recorded in terms of the number of correct anticipations (i.e., the number of correct response words spoken before the shutter raised) on each trial.

Results

The mean numbers of correct anticipations for the six trials on the practice list, 10 trials on the warm-up list (list A), and 10 trials on the test list (list B), are shown in Table I, columns 1, 2, and 3, respectively, for the separate experimental groups. Examination of columns 1 and 2 of this table reveals no substantial differences in this index of learning among the different groups of subjects; simple analysis of variance applied to the means of these groups for each of the two measures involved yields *F*-ratios less than one, indicating that the small observed differences cannot be considered significant. There is no evidence that up to the point of the intro-

TABLE I

Mean Number of Correct Anticipations During Learning of Practice List, Warm-up List (List A), and Test List (List B)

Condition (Interval Between Lists A and B)	Practice (6 trials)	List A (10 trials)	List B (10 trials)
8 Sec.	16.76 ± 2.37	34.53 ± 4.05	75.92 ± 4.99
5 Min.	16.00 ± 2.23	32.92 ± 4.49	70.40 ± 4.51
10 Min.	17.88 ± 1.92	35.04 ± 3.91	72.72 ± 4.23
20 Min.	16.80 ± 1.69	34.96 ± 3.92	65.20 ± 5.02
40 Min.	16.24 ± 1.87	35.40 ± 3.57	55.00 ± 5.71
60 Min.	17.56 ± 1.99	35.08 ± 4.11	54.24 ± 5.15
120 Min.	16.32 ± 1.93	36.12 ± 3.82	60.52 ± 4.71
180 Min.	17.76 ± 2.98	33.64 ± 2.88	51.30 ± 4.23
240 Min.	17.96 ± 1.83	30.32 ± 3.48	48.38 ± 3.99

duction of different lengths of rest interval (between the warm-up list and test list) for the various conditions that the amounts of learning exhibited by the experimental groups were in any way different. Column 3 of this table presents the mean number of correct anticipations per condition for the total of 10 trials on the test list (list B). It may be seen from this column that the mean number of correct anticipations tends to decrease as the length of the preceding rest interval increases. The maximum difference observed is approximately 25 responses and is far greater than any difference observed among the groups on list A prior to any differential treatment. Simple analysis of variance performed on these means yields an F-ratio of 4.01. This value is significant at the one percent level of confidence, an F-ratio of approximately 2.60 being required at this level with 8 and 216 degrees of freedom.

Examination of the data in column 3 of Table I reveals that the successive decrements in learning accompanying increasing length of rest interval tend to become smaller as the length of the rest interval is increased, and that rests of more than one hour appear to have little additional effect on reducing the amount of learning exhibited on list B.

The relationship is indicated graphically in Fig. 1, which shows the mean number of correct anticipations on list B plotted against length of rest interval since previous learning. In this figure the ordinate represents correct anticipations on list B and the abscissa, length of rest; each solid circle represents the mean number of correct anticipations during the ten trials on

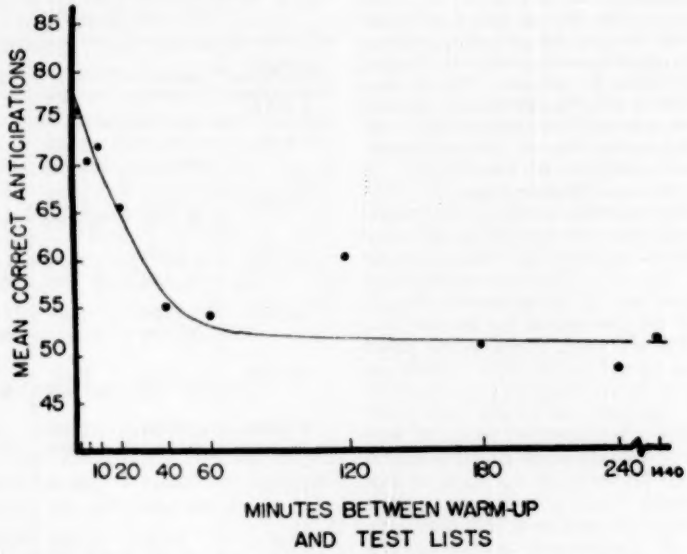


FIG. 1. Mean correct anticipations during 10 trials on the test list following different intervals of rest between test and warm-up lists. Each point represents the mean for a separate experimental group according to the condition of rest introduced. The point above 1440 min. is taken from Thune's data cited.

list B for a particular group of subjects (corresponding to the length of rest received between lists A and B). The value represented by the solid circle above the point labelled 1440 min. (24 hours) on the abscissa was obtained from Thune's data since his materials and experimental conditions for the determination of this particular point were in every respect similar to those used in the present study. This point, therefore, provides a suitable basis for estimating the decrease in learning speed resulting from a 24-hour rest interval and serves to indicate the close approach to the limit of effectiveness of the present experimental conditions. This observation indicates that no abrupt change in the trend determined under the present experimental conditions would appear if a full day intervened between lists A and B.

A smooth curve has been drawn by inspection through the points in Fig. 1 (disregarding for the moment the deviating results of the 120 min. condition). The specific values of the parameters of such a curve, whether drawn by inspection or fitted by some other technique, are not of particular importance to the problem at its present stage of development. The main point of interest represented by this curve is its fairly close correspondence with the data and the fact that it reflects a relatively large initial loss and negatively accelerated relationship between learning decrements and length of time since previous formal learning. The obvious instances of departure of the obtained means from the fitted curve probably reflect the relatively large variability in the measure of performance used. For example, although the mean for condition employing 120 min. rest appears to indicate a sharp reversal in the trend of the data, the differences between this mean on list B and the means for adjacent conditions employing 40, 60, 180, and 240 min. are not shown to be significant on the basis of a *t*-test, a difference of at least 13.20 being required for significance at the five percent level.[3]

The preceding analysis indicates that an over-all differential exists among the experimental groups following different lengths of rest interval. The measure employed (total number of correct anticipations during 10 trials on the test list) does not indicate where during this learning period such differences may be at a maximum following their establishment. Thune (5) presents evidence from his experimental situation in which a similar differential was established, that the exhibited differences in learning speed are maximal during the initial portions of the new learning and tend to become smaller with additional learning trials. A parallel finding might be expected in the present experiment. The same type of analysis as employed by Thune is used in the present instance. For purposes of such analysis the learning curves for list B were divided into thirds, each third including three anticipation trials. The average slope within each third was measured in terms of the mean increase in the number of correct anticipations during that third of the learning. The results of this analysis are shown in Fig. 2. In this figure the mean increase in number of correct anticipations is plotted against each of the three divisions of the learning of the test list (trials 1–4, 4–7, and 7–10). It may be seen that the greatest differences among the conditions in mean increase in correct anticipations during the first third of the new learn-

[3] The *t* for this comparison was based on the within-group variance estimate derived in analysis of variance of these means.

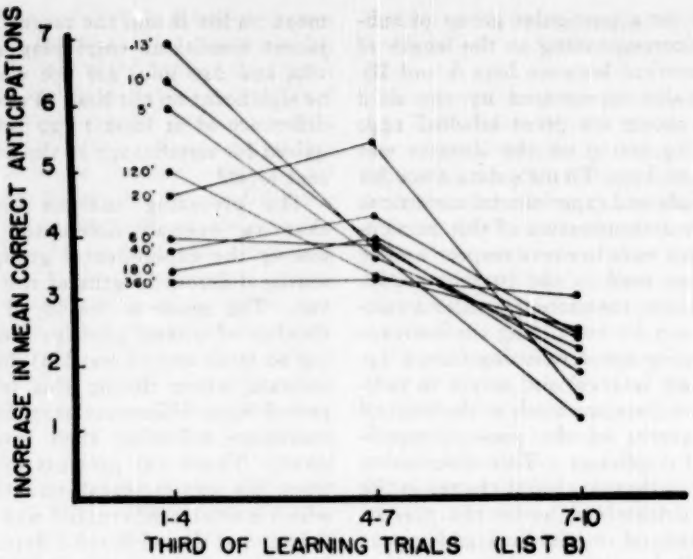


FIG. 2. Mean increase in correct anticipations during each third of the learning of the test list for the nine experimental conditions. The notations for each curve indicate the length of rest separating the learning of the warm-up list and the test list for that condition.

ing; these differences tend to become smaller during the second third and still smaller during the final third. This figure strongly suggests that the retarding effect on the new learning of the preceding rest interval acts primarily during the first few trials in the learning of the new list.

The reversals seen among the curves during the second and final thirds of the learning probably occur because learning under conditions exhibiting the greatest mean increase during the first third tend to approach a ceiling of performance beyond which little additional gains are possible during the final third. Conversely, the conditions exhibiting the smallest mean increase during the initial third tend to exhibit correspondingly larger mean increases during subsequent thirds as they approach nearly the same limit. Some of the conditions show greater mean increases during the second third of the learning than during the first third; this is due to the tendency towards initial positive acceleration during the learning on list B for these conditions.

DISCUSSION

The experimental operations defining warm-up effects that are contained in a general manner within this study are somewhat different from those gained from the earlier work of Bell (2), Ammons (1), and more recently, Irion (3, 4). These investigators were concerned with the relearning situation and their experimental operations were set in such a framework. Insofar as they were concerned with the warm-up variable the definitions proposed were in terms of the slope of the initial segments of the relearning curves without application to more general situations (although Irion has stated the possibility). The

study by Thune serves to promote a more general set of operations applicable to the problem; e.g., ". . . in terms of the slope of the initial segment of a corresponding learning *or relearning* curve obtained after a different amount of practice and/or a different length of retention interval" (5). Such redefinition permits the identification of warm-up effects in the situation involving the learning of new verbal materials as well as those involving the relearning of previously practiced material. Insofar as similar results are observed, additional support is lent to the validity of findings in the earlier situations.

In the present experiment, of course, the operations defining warm-up are even more general than those pertaining to slopes of some particular segment of the learning curves. It was assumed, simply, that if different groups of subjects came to the test list following different amounts of rest interval since previous formal learning, they would do so under conditions of different amounts of warm-up to the situation (i.e., in some inverse proportion to the length of the rest intervals employed). Since the operation of other variables which might affect the rate of learning of the test list was reduced to a minimum, residual differences among the groups in exhibited learning (without confining observation to a specific part of the learning curve) on the test list should then be attributable to differences in state of warm-up at the time this learning was begun.

In the first part of the analysis of data presented, the differences among the experimental groups in terms of the total correct responses during the learning of the test list were demonstrated to be significant. These differences were obtained through employing rest intervals within the limits already available from Thune's data (i.e., 24 hours) and serve to indicate that losses of warm-up gained during learning appear to occur rapidly with the cessation of formal practice and less rapidly as additional rest ensues. Within the conditions of the present experiment, such losses appeared to have reached their maximum by approximately one hour. Some simple negatively accelerated function appears to be suitable to adequately describe this change in time. Repeated duplication of the design using a variety of learning materials might eventually lead to some consistent 'best fit,' which could be offered with more confidence than the simple inspection curve presented, but no rational grounds exist *a priori* to expect one particular function to the exclusion of others at present. In any event the type of function discovered by the best-fitting technique would be the product of somewhat arbitrary decisions as to whether the measure of learning differences should be single trials, tangential points on a learning curve arbitrarily decided upon, or, possibly, arbitrarily selected groups of trials (as in the present experiment). For these reasons, stopping somewhat short of a completely exhaustive interpretation of the data at hand seems most consistent with their refinement and the demands of the problem.

In the second part of the preceding analysis of data that bear on the locus within the learning of the test list for the maximum effects of a warm-up differential, it was seen that the initial trials on the test list showed the greatest differences among the various conditions. This finding may be taken as indicating that the initial differences in state of warm-up among the groups were lost as practice trials on the test list ensued and the groups having had the longest rest intervals

(suffering, presumably, the greatest losses in warm-up adjustments acquired during previous learning) regained these adjustments rapidly with additional practice on the test list. This is consistent with Thune's findings on a similar analysis. It should be noted that whether the subjects begin the test list following a constant amount of immediately prior practice on the warm-up list and different amounts of intervening rest, or different amounts of immediately prior practice and a constant rest period,the problem at the time of the test list is exactly the same—different levels of the state of warm-up presumably exist for the different groups of subjects and the locus of maximum differences in learning of the test list (reflecting the hypothetical warm-up differential) should be observed to be the same; this has been the case with these concurrent studies.

The results of this and related studies demonstrate that a factor of warm-up can operate under certain conditions to produce differences in performance in learning situations. Two parameters of this factor have been investigated so far: (1) its acquisition as a direct function of massed practice on a learning task, and (2) its loss following acquisition as an inverse function of time of rest following formal practice. Evidence presented by Thune further indicates that the acquisition of such a state does not depend on actual learning taking place in the learning situation, but only that subjects 'go through the motions' which are necessary in the actual learning. The precise nature of this factor is not of primary importance except that as it is adequately described, it may be differentiated from other factors affecting performance and its effect controlled where necessary. It would be premature to suggest the necessity of such explicit control in all experimental situations where state of warm-up might operate in a variable manner. This is for the reason that the acquisition of the appropriate attentive and muscular adjustments termed warm-up may itself be learned in such a manner that the effects observed in the present experiment employing naive, relatively unpracticed subjects might fail to appear were highly practiced subjects used. In other words, the crux of the issue is the *rate* at which a warm-up state is achieved by the subjects during learning. At least under the conditions of the present experiment, such a state was probably regained slowly enough at the time of the test list for differences in this process among the conditions to be reflected by significantly different performances. Were the experimental design such that subjects would achieve a high degree of practice prior to completion of the experiment (as with a counter-balanced design, for example, in which each subject serves in all the conditions), no such differences might appear. This possibility offers subject-matter for further investigation to gain the desired answer.

Summary

1. The present study was designed to ascertain the nature of the relationship between level of performance on a new list of paired-associates and the length of interval separating this list from a previously learned list when practice on the former list is held constant. It was expected on the basis of other studies that the relationship between length of interval and the new performance should be inverse; as the interval of rest is increased, performance level on the new list should decrease.

2. Nine intervals of rest ranging

from eight sec. to four hours were employed to test this relationship. Each interval of rest comprised a separate experimental condition. Individual groups of 25 *Ss* were utilized in the study of each condition; a total of 225 *Ss* were employed. The learning materials were lists of paired adjectives or nouns.

3. Records were obtained of the number of correct anticipations during ten trials on the new list under these conditions. As the period of rest separating this list from the preceding list was increased, significant decreases in performance on the new list were observed. The observed decrease appeared to approach a maximum with intervals of approximately one hour. A negatively accelerated function was offered as best describing the observed relationship. It was proposed that the results of the present experiment can best be interpreted in terms of response set and the conditions under which it is produced and lost. This response set fact has been termed *warm-up effect.*

4. Analysis of the mean increase in correct anticipations during each third of the learning trials on the new list indicated that the differences in performance among the conditions were at a maximum during the first few learning trials.

5. It was suggested that the rate at which subjects acquire warm-up set may itself be modified with practice. This possibility is offered for further investigation.

(Manuscript received
September 5, 1949)

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, 54, 263–281.
2. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, 54, No. 243.
3. Irion, A. L. Reminiscence and retention as a function of length of rest and amount of pre-rest practice. Unpublished Ph.D. Thesis, State Univ. of Iowa, 1947.
4. Irion, A. L. The relation of 'set' to retention. *Psychol. Rev.*, 1948, 55, 336–341.
5. Thune, L. E. An investigation of the effects in humans of different amounts and types of preliminary activities on subsequent learning of paired associate material. *J. exp. Psychol.*, 1950, 40. (In press)

ABSENCE OF REMINISCENCE IN THE SERIAL ROTE LEARNING OF ADJECTIVES

BY CLYDE E. NOBLE[1]

Tulane University

INTRODUCTION

Contrary to the classical obliviscence curves of Ebbinghaus (8), several psychological investigations (1, 10, 11, 28, 29) dealing with human learning and retention have reported characteristic initial increases in the recall of partially learned materials following periods of no practice (cf. also 2, 19, 31).[2] This class of phenomena was designated *reminiscence* by Ballard (1). The present paper, however, is concerned only with the short-time increment observed in serial verbal learning by the anticipation method, known as the Ward-Hovland effect. Ward (28) and later Hovland (10, 11), developed a procedure which defined reminiscence as the superior recall performance of an experimental group given a non-practice rest period over that of a control group without rest. Hovland obtained reminiscence following one presentation of a list of 12 nonsense syllables and also after learning to a $\frac{7}{12}$ criterion. Hull and his associates (15) have accounted for this phenomenon in terms of the unequal relative decay of excitatory (E) and inhibitory (I) processes and the consequent spontaneous recovery of effective reaction potential ($\bar{E}$) during rest. Their formal prediction (Theorem XXXV, Corollary 1) is as follows:

> In rote series learned by massed practice to the point where some, but not all, of the syllables are recalled correctly, the mean number of syllables susceptible of correct recall will increase for a certain length of time after the termination of active learning (15, p. 257).

Experimental evidence (10, 11, 15, 28) indicates that the optimal postlearning rest interval for this effect, in the case of nonsense syllables, is approximately two min.

It could be argued that this deduction is applicable also to the serial learning of 'sense' or 'meaningful' materials but there is evidence to the contrary. Deese (7) using poetry and the method of complete presentation, and Melton and Stone (20) using dissyllabic adjectives, 16-item lists, and more stringent criteria of mastery than those of Ward or of Hovland failed to confirm the implicate. Reminiscence also failed to appear in the latter investigation at the $\frac{13}{16}$ level of mastery when the exposure rate was increased from a two-sec. one to one of 1.45-sec., a condition which theoretically would favor its occurrence. Melton and Stone (20) and Buxton (2) have emphasized the distinction between (a) an increment in recall following rest and (b) the effectiveness of distributed practice upon relearning. Melton and Stone found no evidence for the former, but there were insignificant indications of the latter for the two-sec. rate and for the 1.45-sec. rate.

Earlier, Shipley (25) had reported nonreminiscence when lists of 8, 14, and 20 nonsense syllables were presented at a 2.3-sec. rate, but his failure to find reminiscence may have been a function of several methodological variations upon the Ward-Hovland design: (a) his interlist interval was 9.2 sec., (b) presentation was by movie projection, (c) his *S*s read jokes or worked puzzles during rest periods, and (d) the criteria of partial learning were different.

Stone (26), using lists of 12 nonsense syllables exposed at a two-sec. rate and learned to a $\frac{7}{12}$ criterion, tested for reminiscence after intervals of 2, 5, and 20 min. He found no increment in recall after two min. but significant decrements

[1] Now at the State University of Iowa.

[2] Similar results have been reported with conditioned responses by Pavlov (21) and by Calvin (6), and with perceptual-motor learning tasks (cf. 2), which is more than an analogy according to the Lepley hypothesis (16, 17).

after 5 and 20 min. *There was, however, an insignificant increase in the number of Ss showing recall gains after two minutes, with highly significant decreases later.* Stone's analysis of the *S*s who reminisced pertained only to the percentage difference between the zero- and the two-min. conditions. The writer has computed χ^2 values for each of the frequency distributions exhibited by Stone under the 0, 2, 5, and 20 min. delay conditions, finding marginal forgetting ($P > 0.05$), no change ($P = 1.00$), forgetting ($P < 0.01$), and forgetting ($P < 0.001$), respectively. These tests are the basis of the statements given in italics, for which the present author assumes responsibility. It is pertinent to note here that these absolute differences as functions of time are compatible with Hull's theory except for the non-gains in recall at two min. Stone suggested that the high variability of his data lessens the reliability of his conclusions.

Recently, Buxton (3) reported a study of both nonsense syllables and adjective material, in which 16 items were presented at a two-sec. rate with a six-sec. interlist interval, but learned to a prerest *mastery* criterion. He noted an insignificant gain for adjectives after 0.5 min., then losses after 2, 5, 10, and 20 min. With the syllables there was a significant initial increase up to two min., then decreases thereafter, in accordance with theoretical expectation.

Psychologists studying reminiscence have found the problem of controlling *S*'s activities between learning and recall a vexing one, especially the matter of rehearsal. Ward (28) had his *S*s read jokes or simply give color-associations during the rest period, finding that reminiscence was greater under the latter conditions. Hovland's (10, 11) *S*s named colors exposed at a two-sec. rate, as in the present study. The *S*s of Melton and Stone were instructed to name colors from cards 'as fast as possible' (20, p. 296); and Buxton (3) gave his *S*s 'light magazines' to read. It is probably impossible to eliminate rehearsal entirely. Ten *S*s in the present experiment reported intrusions during rest or during the interlist interval, and other investigators (4, 5, 30) have encountered deliberate rehearsal with their *S*s. However, Buxton and Ross (5) have shown that neither the rehearsal variable nor that of the learning method (pronunciation *versus* spelling the syllables) is of crucial importance in determining reminiscence. Withey, Buxton, and Elkin (30) have found, furthermore, that the nature of the color-naming activity does not produce significantly different recall scores, nor does the rate of naming seem to be important. Apparently we must search for other variables to account for the differences between the Ward-Hovland and the Melton-Stone results. A possibly crucial factor is the variation in stimulus-meaning occurring in these experiments (2).

In order to contribute further to the research on this problem the present experiment was carried out. The procedure corresponds more closely to that adopted by Hovland (10, 11) than did the studies of Deese, Shipley, Melton and Stone, Stone, or of Buxton, but it does not duplicate Hovland's conditions.[3] The design is comparable in all essential respects to those portions of his investigations dealing with massed practice effects, the principal altered variable being the degree of meaningfulness of the rote materials.

Procedure

Apparatus.—Two memory drums of a modified Lipmann synchronous motor type, Tulane design, were used to expose the adjective lists. The size of the aperture was 2 × ¾ in. and the band width was three in. Six bands of adjec-

[3] Both Ward (28) and Hovland (10, 11) used *PE*'s as measures of variability then carried out tests of significance with the CR_{PE}. According to Peters and Van Voorhis (22) this is questionable practice and should not be applied to empirical distributions. Analyses in the present experiment are based on small sample theory, and all but one (cf. Table I) utilize the formula for the significance of differences between the means of matched pairs.

tives and one band of colors were prepared. The three practice and three experimental bands consisted of lists of 12 trisyllabic adjectives, each list beginning with a cue symbol formed by a row of asterisks. The words were neither synonymous nor antonymous, and none appeared in more than one list. They were selected from the middle range (4000 to 6618) of Thorndike's (27) list of the 10,000 most frequently used words. No word had the same initial letter or syllable in any list, and not more than three in a single list had the same terminal letter or syllable. Each adjective began with a consonant syllable, the order in each list being randomized. All possible combinations of situation order (see below) were used, and the order of lists was assigned to these combinations at random. *Ss* were assigned to the several situations by the same method. The adjectives were typed in 14-point black capitals. The color band, consisting of two consecutive series of 16 different hues occurring in random order, was prepared by applying crayon directly to the paper. All randomization was effected by means of the table of random numbers given in Lindquist (18).

The drums were placed side by side on a table facing *S* with adequate illumination from behind. *E* sat to *S*'s right, where he controlled the administration of stimuli and recorded *S*'s responses. Experimentation was carried out in a small room in which auditory distractions were minimized. The presentation rate of both adjectives and of colors was two sec., and the interlist interval was six sec.

Method.—Twenty-four male undergraduates, naive to memory experimentation and uninformed of the purpose of the research, were engaged on a volunteer basis. They served as *Ss* in a serial rote learning situation using the method of anticipation with massed practice. Each *S* reported to the laboratory for six consecutive days, of which the first three were devoted to practice and the last three to experimental sessions. The instructions were modified from Hull *et al.* (15) and included the requirement that *S* correct erroneous anticipations. All experimental procedures were given during the practice sessions, and the situation for any given day was one of the following: I. *One presentation of the list, rest pause, recall after two min.* Immediately after the last word was exposed, the color drum was started and *S* began naming the colors as they appeared. After two min. this drum was stopped and *S* attempted to recall the words by anticipating them on the adjective drum. II. *Learning to a criterion of 7 out of 12 words correct, rest pause, continuation of learning to mastery (defined as one perfect recitation).* On the trial immediately following *S*'s correct anticipation of seven or more adjectives, color-naming was begun as in Situation I. After two min. *S* resumed learning until mastery. III. *Learning to mastery.* In this procedure *S* continued learning by massed practice until reaching one perfect recitation.

The following scoring criteria were adopted: a response was judged incorrect if it was preceded by two successive identical anticipatory errors, or if the adjective was mispronounced or only partially enunciated, or if an otherwise correct response coincided with the appearance of the response-word in the aperture of the memory drum.

Results

Lack of evidence for reminiscence.—Table I presents the results of a comparison of Situation I with Situations II and III. There is no evidence for a difference in retention following one presentation of the list with and without the two-min. rest ($P > 0.80$). The analysis of individual *Ss* showing gains or losses with time, in fact, suggests forgetting since 58 percent of them suffered a decrement in retention. In order more adequately to evaluate these data, McNemar's χ^2 test of the significance of changes was carried out. This procedure tests the null hypothesis that equal frequencies of *Ss* will show increments and decrements in retention ($N = 24$, $df = 1$).

TABLE I

Retention Following One Presentation of List with and without Two-Min. Rest between Presentation and Recall

Group Analysis

	Mean	σ
No. adj. correctly recalled immediately:*	2.14	1.46
No. adj. correctly recalled after two min.:	1.67	1.20

$D = 0.47$; $t = 0.25$; $P > 0.80$.

Individual Analysis

No. *Ss* showing recall increment:	5
No. *Ss* showing no change in recall:	5
No. *Ss* showing recall decrement:	14

$\chi^2 = 3.37$; $P > 0.05$.

* In this case the data for Situations II and III were pooled, necessitating the use of the formula for independent samples.

TABLE II

COMPARABILITY OF EXPERIMENTAL AND CONTROL GROUPS BEFORE INTRODUCTION OF REST PAUSE

Condition	Mean No. Adj. Correctly Antic. on Crit. Trial	Mean No. Trials to Reach Crit.
Without rest pause (Sit. III)	7.58	5.71
With rest pause (Sit. II)	7.79	5.46
P from t-test of difference	>.30	>.70

After applying Yates' correction for continuity, this distribution was found to be of borderline significance ($P > 0.05$) in its departure from chance expectation. There is, therefore, no evidence for reminiscence under this condition.

Table II shows the comparability of the two groups in Situations II and III before the introduction of the rest pause. It will be noticed that, by two independent measures, the groups are equated since the differences in the mean number of correct anticipations on the criterion trial and in the mean trials required to reach that criterion are not significant ($P > 0.30$, $P > 0.70$, respectively).

The principal results of the present experiment are exhibited in Table III. The recall scores (III-A) show a difference in the direction of a decrement in retention following rest when learning has been to the criterion of seven words correct, but the difference is not significant ($P > 0.30$). As in Table I, forgetting is again suggested by the individual analysis, with 54 percent of the *Ss* showing a decrement; but the frequency data yield a χ^2 which is not significant ($P > 0.10$). When the effects of distributed practice are indicated by the number of trials required to relearn to the 7/12 criterion (III-B), there is no difference ($P > 0.80$). Finally (III-C), when the trials required to proceed from seven correct to mastery are computed for the rest and nonrest conditions the difference is not quite significant ($P > 0.05$), although it is in the direction expected on the basis of other empirical findings (cf. **2**, **19**, **31**).

In Fig. 1 the composite performance curves for the two conditions of the experiment are plotted. The correct anticipations are shown as a function of the number of trials. It is evident that the introduction of the two-min. rest pause following attainment of a 7/12 criterion resulted in accelerated learning. Both curves assume the form of the positive growth function deducible from the postulates of Hull's (**14**) general theory of behavior.

Serial position effects.—The graphs in Figs. 2 and 3 show the type of serial position effects deduced by Hull and

TABLE III

RETENTION WITH AND WITHOUT TWO-MIN. REST PAUSE FOLLOWING LEARNING TO CRITERION OF 7 OUT OF 12 ADJECTIVES CORRECT

(A) Recall

Group Analysis

Mean no. adj. recalled immediately after crit. trial:	7.33
Mean no. adj. recalled 2 min. after crit. trial:	6.96

$D = 0.37$; $t = 0.95$; $P > 0.30$.

Individual Analysis

No. *Ss* showing recall increment:	6
No. *Ss* showing no change in recall:	5
No. *Ss* showing recall decrement:	13

$\chi^2 = 1.89$; $P > 0.10$.

(B) Relearning to 7 Correct

Mean no. trials required to relearn without rest:	1.46
Mean no. trials required to relearn with 2 min. rest:	1.50

$D = 0.04$; $t = 0.20$; $P > 0.80$.

(C) Proceeding From a Criterion of 7 Correct to Mastery

Mean no. trials to mastery without rest:	9.87
Mean no. trials to mastery with two-min. rest:	7.87

$D = 2.00$; $t = 1.80$; $P > 0.05$.

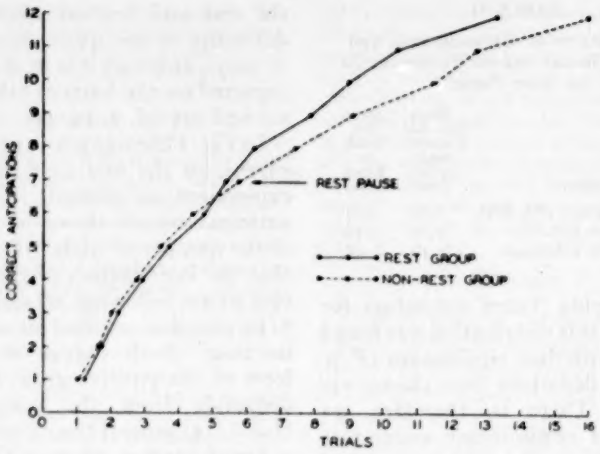


FIG. 1. Performance curves of mean number of repetitions of list required to anticipate correctly varying numbers of adjectives with and without two-min. rest pause (color-naming) introduced after reaching criterion of 7 out of 12 correct by massed practice. Each point is the mean of 24 determinations.

his associates (15) and validated by Ebbinghaus (8), Ward (28), and Hovland (10, 11) for nonsense syllables, by Robinson and Brown (24) for three-place numbers, and by Melton and Stone (20) for dissyllabic adjectives. Fig. 2 presents the mean failures of correct anticipation as a function of the ordinal position of adjectives in the lists, when learning has progressed to the criterion of seven words correct. Fig. 3 shows the mean failures in various serial positions occurring between 7/12 and 12/12 correct, with and without the two-min. rest. This graph illustrates the advantages in error-savings of distributed practice in the central portion of the series. It confirms Theorem XXI, Corollary 3 (15), which already has been verified for nonsense syllables by Hovland (10, 11, 12, 13) and by Patten (23). It is to be noted that Melton and Stone's curves (20, Fig. 3) show increasing serial errors with longer rest intervals (2, 5, and 20 min.). This has been cited (2, 20) as evidence of a reversal of the relative reminiscence effect predicted for the central posi-

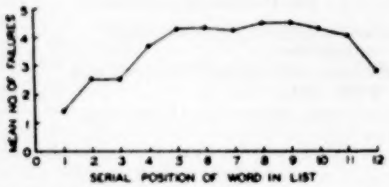


FIG. 2. Serial position effects: mean number of failures during learning to criterion of 7 out of 12 adjectives correct. Data represent values from Situation II, each point being the mean of 24 determinations.

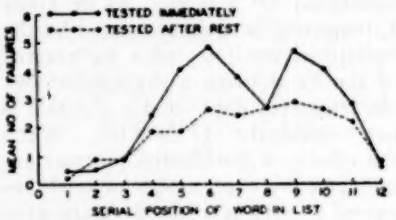


FIG. 3. Serial position effects: mean number of failures in various positions between criterion of seven correct and mastery, with and without two-min. rest pause. Each point is the mean of 24 determinations.

tions. This evidence must be qualified, however, by stating the operations of measurement. Melton and Stone made their observations on the first postrest trial and not, as in the present experiment, throughout the postrest relearning period. Their data do not, therefore, falsify this deduction.

Theorem XXII, Corollary 1, referring to the greater mean number of failures at the posterior position than at the anterior position, is verified by the curve shown in Fig. 2. Theorem XXII, Corollary 3, however, is probably not confirmed, since the curves in Fig. 3 fail to show a greater economy of distributed practice at the posterior position. Theorem XXXI, Corollary 3, is not confirmed by this experiment. This corollary states that the eighth word in the series should show the maximum superiority of distributed practice over massed practice. The data in Fig. 3 indicate, to the contrary, that positions 6 and 9 profit most by the interpolated rest, position 8 showing no change. The dip in the serial position curve for the nonrest condition at positions 7 and 8 is unaccounted for, although it may be an artifact of the heterogeneity of the lists. The method used in constructing them may certainly be improved upon, and a standard procedure for calibrating meaningful lists is needed in research of this type.

Discussion

There is no evidence for reminiscence, either following one presentation of the list or when learning has proceeded to a 7/12 criterion. The absolute differences measured by recall scores and by the number of *Ss* showing changes are, moreover, in a direction opposite to that found by Ward and by Hovland. Similarly, there is no indication of the effectiveness of interpolated rest in the case of relearning to a criterion of seven words correct. But there is an insignificant saving in the number of trials required to proceed from this criterion to mastery following rest. Thus, a dilemma presents itself: why does the reminiscence phenomenon not occur under conditions regarded as optimal (for nonsense syllables) when the major condition varied is along the dimension of stimulus-meaning? Some have pointed out that reminiscence designs have not been adequately standardized to permit the identification of such a critical variable. The objections raised, e.g., by Melton and Stone (**20**) and by Buxton (**2**) in regard to such inter-experimental differences as spelling *vs.* pronouncing, varied rest interval activities, and correction *vs.* non-correction are probably now of little significance in the light of subsequent research, particularly the work of Buxton and his associates (**4, 5, 30**). The present design more nearly satisfies the Ward-Hovland criteria for producing reminiscence than have other recent experiments, and may therefore be a more valid test of the phenomenon than are those in which variables known to be relevant (length of list, criteria of partial learning and of mastery, practice effects) have not been properly controlled.

The failure of reminiscence to occur in the learning of trisyllabic adjectives, chosen for their formal similarity to nonsense syllables, and under standardized conditions with respect to number of items, degree of acquisition, test-retest controls, presentation rate, interlist interval, length of rest, color-naming activity, set, learning technique, and level of practice is a finding of some consequence to learning theory. These results are also consistent

with the results of several analogous investigations (3, 7, 20).

Serial position effects occurring in this experiment are in general agreement with the predictions of Hull *et al.* This constitutes evidence in support of their mathematico-deductive theory of rote learning with respect to the concept of intraserial inhibition and the conditions of its removal.[4] It suggests further that their theoretical system, although designed for nonsense material, may with certain modifications account for the phenomena of serial verbal learning, including the use of 'sense' lists. The writer believes it is possible to account for non-reminiscence and obliviscence under these conditions in terms of decreased habit interference effects at the termination of active learning. It may be postulated that the increments of interference (ΔI) are greater during the learning of nonsense syllables than in learning meaningful items. This assumption is based on the positive empirical relationship found by several investigators between meaningfulness and rate of learning. The studies reviewed by McGeoch (19) suggest that this function is linear within definable limits. Accordingly, it is supposed as a first approximation that the magnitude of these serial increments varies inversely as some function of the degree of meaningfulness (m) of the rote material; i.e., $\Delta I = f\left(\frac{1}{m}\right)$. A consideration of this and certain other relevant variables yields a functional equation regarding the growth of I:

$$I = f\left(n, s, \frac{1}{m}, \frac{1}{t}\right),$$

where $n \equiv$ number of stimulus items in list; $s \equiv$ index of physical similarity of stimulus items; $m \equiv$ index of meaningfulness of items, such that $0.00 \leq m \leq 1.00$; and where $t \equiv$ duration of inter-item interval in seconds.

It is thus postulated that familiar and meaningful material is easier to learn than unfamiliar or meaningless material because it has associated with it less I. Furthermore, material which generates less I from other sources—such as n, s, and t—will also be less difficult to learn. With the cessation of practice and serial reinforcement, the value of effective excitatory potential ($\bar{E}$), where $\bar{E} \equiv E - I$, will decrease steadily with time rather than undergo an initial increase. This state of affairs is then observed behaviorally as unaltered recall or as forgetting depending, other things constant, upon the empirical value assigned to m.[5] Such a deduction is in conformity both with the present results and with the theoretical derivation of the Ward-Hovland phenomenon by the Yale group. The present point of view simply regards reminiscence as a special case of a more general principle of retention.

[4] Certain psychologists have contended that reminiscence experiments have not conclusively demonstrated the dependence of the phenomenon upon the dissipation of inhibition. This is to question whether reminiscence is consistently associated with certain operationally specified indices of inhibition (e.g., center serial position effects). Although Melton and Stone's discussion indicated that it is not, it should be pointed out that the evidence they reported (20, Fig. 1-A) for an increase in intraserial inhibition with increased rate of presentation was not statistically significant. Neither, for that matter, are the differences between the two serial position curves shown above in Fig. 3 ($P > 0.05$), yet both the Melton-Stone data and those of the present author are suggestive of *small* accumulations of habit interference. It is implausible to expect the development of I to be an all-or-none affair. Hence, one wonders whether Melton and Stone's results are not, in fact, consistent with the theory considered in this paper.

[5] Available evidence indicates that under standard conditions reminiscence will appear optimally when the value of $m < 0.30$.

The postulate set designed by Gibson (9), based on the concepts of conditioned generalization and differentiation, may account for additional mechanisms underlying these retention phenomena. She regards such interference effects as those found in nonsense lists as the result of intraserial generalization. Differences in recall or in work-savings are assumed to be functions of the extent of initial stimulus generalization. Further, generalization tends toward a minimum for meaningful items and toward a maximum for nonsense items. When it is observed, therefore, that more trials are required to learn nonsense syllables, the explanation is in terms of the difficulty of establishing conditioned differentiation among the several items. In other words, when meaning is defined as the differentiating characteristics of stimuli one from another, ease of learning is proportional to degree of meaningfulness. This definition of stimulus-meaning is useful but somewhat equivocal, since intralist similarity-difference is a dimension logically distinct from that of meaning-nonmeaning. Thus, material may be highly meaningful and acoustically similar (e.g., noun homophones), or relatively meaningless and visually similar (e.g., homomorphous nonsense syllables of low association value), yet the behavioral effects of both would, in Gibson's view, presumably be referable to the same set of explanatory principles. In view of the *conceptual* independence of these dimensions, therefore, it seems reasonable to use an association value as the index of stimulus-meaning (m) and to define intralist similarity (s) in terms of the physical (formal) characteristics of the stimuli (e.g., by a psychometric scaling method). Accordingly, the functional equation for I above has included expressions to denote the effects of variation along each of these dimensions.

Certain alterations seem to be required in Hull's system if it is to account for serial verbal learning phenomena other than those for which it was designed. The changes suggested here are only tentative; more precise formulations await further empirical data.

Summary

1. Twenty-four practiced *Ss*, equated as to pre-rest level of acquisition, learned lists of 12 trisyllabic adjectives by the anticipation method at a two-sec. presentation rate with a six-sec. interlist interval. Recall and relearning data computed with and without a two-min. rest interval of paced color-naming were analyzed for reminiscence effects.

2. There was no evidence for reminiscence following a single presentation of the list. An analysis of the performances of individual *Ss* indicated forgetting of marginal significance.

3. The introduction of a two-min. rest following learning to a criterion of seven words correct resulted in no difference in the mean number of words correctly anticipated on the first postrest trial, nor did it yield a work-saving in relearning to the same criterion. This was the crucial test for reminiscence in the present experiment. There was, however, a saving of borderline significance in the mean number of trials required to continue learning to mastery from that criterion following rest.

4. The effect of interpolated rest, considered graphically in composite curves, was to reduce the negative acceleration of the performance growth function. The error-saving, in terms of serial position effects, appeared most clearly in the central and postcentral portions of the lists.

5. The present findings with respect to serial position effects are in partial conformity with those of Ward and of Hovland, but they fail to confirm their reminiscence effects as defined by differential recall. The data are in general agreement with those of Melton and Stone and of Buxton.

6. These results seem adequately accounted for in terms of the differential decay of inhibitory and excitatory processes, according to the theoretical analysis of Hull *et al.* Certain modifications seem necessary in their system, however, if it is to mediate the deduction of serial verbal learning phenomena using meaningful items as well as nonsense syllables.

7. An alternative postulate is proposed for the growth of I which utilizes the mean association value, m, of the material as an index of stimulus-meaning. It is suggested that this is a critical variable operating in recent experiments on reminiscence.

(Manuscript received
September 14, 1949)

References

1. Ballard, P. B. Obliviscence and reminiscence. *Brit. J. Psychol., Monogr. Supp.*, 1913, 1, No. 2.
2. Buxton, C. E. The status of research in reminiscence. *Psychol. Bull.*, 1943, **40**, 313–340.
3. Buxton, C. E. Interaction of meaningfulness and length of rest interval as factors in measured reminiscence. *Amer. Psychologist*, 1948, **3**, 235–236.
4. Buxton, C. E., & Bakan, M. B. Correction vs. non-correction learning techniques as related to reminiscence in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 338–341.
5. Buxton, C. E., & Ross, H. V. Relationship between reiminiscence and type of learning technique in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 41–46.
6. Calvin, J. S. *Decremental factors in conditioned-response learning.* Ph.D. thesis, Yale Univ., 1939.
7. Deese, E. L. *A study of the retention of meaningful verbal material at short time intervals.* M.A. thesis, Yale Univ., 1938.
8. Ebbinghaus, H. *Über das Gedächtnis.* Leipzig: Duncker & Humbolt, 1885.
9. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
10. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
11. Hovland, C. I. Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
12. Hovland, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
13. Hovland, C. I. Experimental studies in rote-learning theory. VII. Distribution of practice with varying lengths of lists. *J. exp. Psychol.*, 1940, **27**, 271–284.
14. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
15. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.
16. Lepley, W. M. A theory of serial learning and forgetting based upon conditioned reflex principles. *Psychol. Rev.*, 1932, **39**, 279–288.
17. Lepley, W. M. Serial reactions considered as conditioned reactions. *Psychol. Monogr.*, 1934, **46**, No. 205.
18. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
19. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, & Co., 1942.
20. Melton, A. W., & Stone, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.
21. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927.
22. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill Co., 1940.
23. Patten, E. F. The influence of distribution of repetitions on certain rote-learning phenomena. *J. Psychol.*, 1938, **5**, 359–374.
24. Robinson, E. S., & Brown, M. A. Effect of serial position upon memorization. *Amer. J. Psychol.*, 1926, **37**, 538–552.
25. Shipley, W. C. The effect of a short rest pause on retention in rote series of different lengths. *J. gen. Psychol.*, 1939, **21**, 99–117.
26. Stone, G. R. Retention of nonsense syl-

lables over short intervals of time. *Proc. Ind. Acad. Sci.*, 1946, **55**, 178–181.

27. THORNDIKE, E. L. *The teacher's word book.* 2nd ed. New York: Teachers College, Columbia Univ. Press, 1921.
28. WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 4, Whole No. 220.
29. WILLIAMS, O. A study of the phenomenon of reminiscence. *J. exp. Psychol.*, 1926, **9**, 368–387.
30. WITHEY, S., BUXTON, C. E., & ELKIN, A. Control of rest interval activities in experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 173–176.
31. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

TRANSFER OF VERBAL TRAINING TO A MOTOR TASK [1]

BY KATHERINE E. BAKER AND RUTH C. WYLIE

Connecticut College

Introduction

Several investigations in the field of transfer of training have shown positive transfer from so-called mental or 'imaginary' practice to a number of motor tasks. Sackett (5, 6) found that relearning a maze was facilitated by either drawing the pattern or thinking through the maze, with the drawing type of rehearsal having the greater transfer value. In another maze learning experiment (8), the subjects who reported having depended primarily upon explicit or implicit verbalizations learned more efficiently than those who said that they had depended primarily upon visual imagery or a motor mode of attack. Another study (4) compared the efficiency of actual and 'imaginary' practice on five motor tasks, ranging from those demanding simple coordination to those requiring ideational processes. Both types of practice were found to be effective, with the imaginary practice having its maximum effectiveness for those tasks involving ideational activity. In an experiment investigating the relative values of 18 days of physical and mental practice upon dart throwing and basketball shooting (7), it was found that mental practice was almost as effective as actual practice in improving the motor skill.

Although these experiments show amazingly large amounts of transfer resulting from mental practice, none of them provides a basis for analyzing just what kinds of mental practice are occurring and what the theoretical explanation of their transfer value might be. Neither the nature of the task nor the experimental design makes it possible to determine the relative roles of verbalization, visual imagery, or implicit motor responses. In some types of motor tasks, the major difficulty is encountered in learning the relationship between signalling stimuli and required responses, e.g., card sorting, or complex choice reaction time tasks. In such tasks, we can surmise that verbalizations, and perhaps visual imagery, should be the most effective kinds of mental practice. In other tasks, such stimulus-response relationships may be few and simple, and the difficulty of the task lies in learning to associate kinesthetic cues from precise body movements with the visual or other stimuli which signal that the movement has led to correct results, e.g., the correct ball-basket relationship is easy to learn, but it is difficult to discover which kinesthetic cues will be followed by the sight of the ball dropping through the hoop. In this sort of task it is probable that implicit verbalizations would be of little help (except perhaps in reviewing general facts about posture, arm positions, etc.) and that the most effective mental practice would consist of implicit motor responses. In terms of this analysis, many of the tasks used in the experiments cited above were of such a nature that both verbalization and implicit motor practice might have been used with some transfer value, thus preventing one from determining the relative effectiveness of either type of rehearsal. In addition, the design of the experiments involved initial test trials on the actual tasks which were later to be practiced mentally. Thus, the subject had an opportunity to observe the response to be made, and might have practiced them implicitly during the mental rehearsal interval. In such cases, even if verbalization was the chief type of mental practice used, its effect would not be clearly separable as such.

Another factor which has not been adequately controlled in previous ex-

[1] The money used to pay for the services of experimenters was supplied under Contract N7onr-316, Task Order I, between Special Devices Center, Office of Naval Research, and Connecticut College.

periments is the amount of mental rehearsal given. When attempts are being made to compare types of practice, it would be desirable to have an exact and known number of trials given to each group of subjects.

It was the purpose of the present experiment to control these factors by an experimental design which would markedly reduce the possibility of motor practice occurring simultaneously with verbal practice, and would definitely determine the number of training trials. In the training task, the subject is asked to memorize, by the paired associates method, verbal expressions of the stimulus-response relationships which he will later encounter in a final motor task. The final task was a discrimination learning problem, involving four lights, each of which could be turned off by a given switch. This task seemed to be one of those in which the most difficult associations lent themselves to verbal expressions, so that positive transfer from verbal practice might reasonably be expected here. The possibility of the occurrence of implicit motor practice was further controlled by giving the verbal training without the subject's knowing that any motor task was to follow.

Besides exploring the transfer value of verbal training, we also propose to attempt an explanation of the present results in terms of the theoretical concepts which have emerged from other transfer experiments, using different types of training before this motor task.

Method

The motor task set up for these experiments requires that the subject hit with his hand one of four toggle switches in response to red or green signal lights which may appear in either of two positions on a vertical panel. The subject starts each response from a constant central point and attempts to turn out the presented signal light as quickly as possible by hitting the proper switch on a horizontal panel. Both time of correct response and errors may be used as measures of learning.

The motor task used here is identical with one which has been employed previously in the study of several other aspects of the problem of transfer of training to motor skills. The apparatus is described in detail elsewhere (1).

One group of subjects, a control group, was given no verbal training, but was merely told which one of the switches should be hit when each of the four stimulus lights appeared. This group was then given 60 presentations of the stimulus lights in random order, with the instructions to try to hit the correct switch as quickly as possible after the signal light appeared. Before each presentation of a stimulus light, the subject was given a ready signal which served as a two-sec. warning of the appearance of the light and as a signal to press down with the thumb on a central starting button. An electric timer started with the presentation of a stimulus light and stopped when the subject hit the correct switch. The subject was instructed that if the first switch he tried did not turn out the stimulus light, he was to rectify his error by hitting the correct switch as quickly as possible, since the timer continued to run until the correct response was made. The time of response was recorded for each trial, and the subject was told his time of response as well as being allowed to read the clock for himself. In addition, the number of errors was recorded for each trial, but the subject was not informed that this was being done.

Two additional groups of subjects were given eight and 24 verbal training trials, respectively. This verbal training involved words which expressed the stimulus-response relationships occurring in the motor task. These subjects were informed that they were going to be asked to learn verbally certain relationships which would be used in a later part of the experiment. (No mention was made that the experiment would ever involve a motor task.) They were told that *top-red* went with *right-one*, *top-green* with *right-two*, *bottom-red* with *left-one* and *bottom-green* with *left-two*. During the training series, they were asked to say as quickly as possible the second pair of words which went with any given first pair spoken by the experimenter. The pairs of words presented to the subject were given in random order and he was required to continue responding until he gave a correct answer. In order to motivate the subject for speed the experimenter pretended to time the subject by a stop watch.

Following the verbal training, the subject was taken into the experimental room and shown the motor skill apparatus for the first time. It was

TABLE 1

TRANSFER IN TERMS OF RESPONSE TIME FOLLOWING VARYING AMOUNTS OF VERBAL TRAINING

Means and standard deviations of average response time per trial for successive 10-trial stages in the learning of a motor task are presented for three groups of 18 subjects, each group having had a different amount of verbal training prior to learning the motor task.

	Trials					
	1–10	11–20	21–30	31–40	41–50	51–60
0-Trial Group						
M (.01 sec.)	149.22	118.79	108.07	95.60	100.02	95.24
SD	78.72	55.32	47.08	35.07	54.56	46.94
8-Trial Group						
M (.01 sec.)	151.14	117.81	105.40	107.83	99.08	88.78
SD	85.82	55.17	69.13	49.09	59.26	39.82
24-Trial Group						
M (.01 sec.)	137.66	106.87	97.39	99.86	90.79	87.90
SD	83.31	58.57	45.53	54.42	47.46	40.01

pointed out to him that the words presented to him in the training series represented the color and position of the lights that would appear on the stimulus panel of the apparatus; and that the pairs of response words which he had learned represented the position and number of the switches he was to hit in response to these stimulus lights. The subject was then given 60 trials on the motor task itself, with response time and errors recorded for each trial.

Each group consisted of 18 female college students. The groups were matched on the basis of two criteria: a digit-symbol test with an 80-sec. time limit, and a reaction time measure based on 25 presentations of one of the given lights. These two measures had previously been found to give a multiple correlation of .42 with the average response time for 60 trials on this motor task, and a multiple correlation of .35 with number of errors made in 60 trials on the task (1).[2]

RESULTS

Table I shows the results obtained for the three groups in terms of response time. This table presents the means and standard deviations of response time per trial for each 10-trial stage of learning, and Fig. 1 presents these means graphically. It may be seen that, for all three groups, average response time per trial is high at first, drops rather rapidly with additional trials, and, after a more gradual drop, appears to approach a constant level. The transfer effects of verbal training may be seen by comparing the curves obtained by the experimental groups with that of the control group. Eight trials of verbal training apparently produced no transfer effect, since the curve is practically identical with the control group curve. Differences significant at the .05 level or better appear only during trials 31–40 and 51–60. However, the average response time for the group given 24 verbal training trials is significantly lower than that of the control group except during trials 31–40; and is significantly lower than that of the 8-trial group except during trials 51–60.[3]

[2] On the reaction time test, means for the three groups ranged from .5622 to .5866 sec., while standard deviations ranged from .0649 to .0762. On the digit symbol test, means ranged from 69.00 to 69.39, while standard deviations ranged from 9.83 to 10.95.

[3] The critical ratios used here have been computed by a method which rules out inter-trial variance by computing the standard errors of the differences between the means of each of the 10 trials within each stage, using $N = 10$.

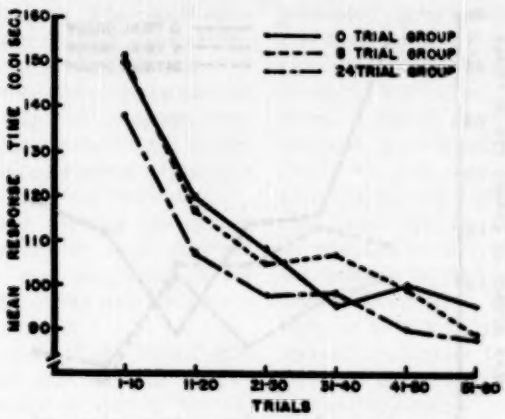


FIG. 1. Learning curves in terms of mean response times per trial for successive 10-trial stages of learning of a motor task. Each curve was obtained for a group of 18 subjects, each group having had a different amount of verbal training prior to learning the motor task.

Table II shows the results obtained for the three groups in terms of errors. Means and standard deviations are presented for the number of errors made during each 10-trial stage of learning. Fig. 2 is a plot of the data of Table II. In general, the two experimental groups make fewer errors than does the control group, the greatest discrepancy appearing during the beginning stages of learning the motor task. For 10-trial stages these differences are not significant, probably because only a few errors were made, and the distributions of these errors were skewed, resulting in large standard deviations. However, when we consider half of the learning trials at a

TABLE II

TRANSFER IN TERMS OF NUMBER OF ERRORS MADE FOLLOWING VARYING AMOUNTS OF VERBAL TRAINING

Means and standard deviations of the errors made during each 10-trial stage and during each half of the learning of a motor task are presented for three groups of 18 subjects, each group having had a different amount of verbal training prior to learning the motor task.

	Trials							
	1–10	11–20	21–30	31–40	41–50	51–60	1–30	31–60
0-Trial Group								
M	2.67	1.67	1.61	1.00	1.56	1.72	6.00	4.28
SD	3.02	1.85	1.95	1.45	1.42	2.21	4.26	4.37
8-Trial Group								
M	1.50	1.60	1.39	1.39	0.72	0.89	4.56	3.06
SD	1.74	1.49	1.34	1.57	0.93	0.94	3.01	2.17
24-Trial Group								
M	1.33	1.17	0.94	1.33	0.83	0.89	3.50	3.06
SD	1.53	1.21	1.23	1.57	0.96	1.41	2.85	3.08

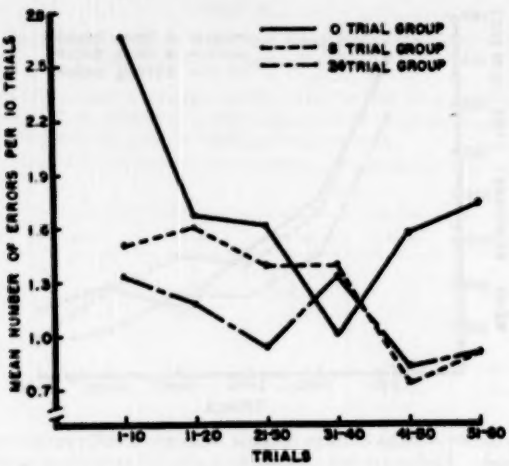


FIG. 2. Learning curves in terms of mean numbers of errors made during successive 10-trial stages in the learning of a motor task. Each curve was obtained for a group of 18 subjects, each group having had a different number of trials of verbal training prior to learning the motor task.

time, a critical ratio of 2.16 is obtained for the difference between average number of errors made in the first 30 learning trials by the 24-trial group and the control group. Thus, 24 trials of verbal training provide some measurable degree of positive transfer to the motor task in terms of errors as well as time. The positive transfer effect, however, cannot be shown to continue after the first 30 trials of learning.

DISCUSSION

This experiment has been concerned with the effects of verbal practice upon the subsequent learning of a motor task. No transfer effect was found following small amounts of training in verbalizing the stimulus-response relationships of the motor task; but larger amounts of such verbal training were of benefit to the later learning of this task. In the present experiment, transfer was obtained when the rehearsal was presumably exclusively verbal. The particular experimental design which was used allowed for restricting the amount of implicit motor practice during training, and the motor task investigated was one which would probably benefit by verbal training. We do not wish to leave the impression that our results and conclusions may be considered to apply to other kinds of tasks in which it would be more difficult to formulate appropriate verbalizations. As was noted in the Introduction, it is difficult to see how certain types of tasks, such as basketball shooting, could be expected to benefit much by verbal training.

Positive transfer resulted here when both the stimuli and responses of the training task differed from those of the final task. A previous experiment (**2**) has shown comparable results for another situation in which training and final tasks involve both different stimuli and responses. In this previous experiment, positive transfer resulted from practice on a pictured representation of the apparatus to a motor task very similar to that of the

present experiment. The panel containing the light stimuli was represented by four printed circles, one of which was colored red or green on any given sheet; and the response consisted of making a pencil mark in one of the four printed squares representing the positions of the switches.

An hypothesis making use of the concepts of stimulus and response generalization was suggested to explain the results of the pictured representation experiment.

In the first part of the hypothesis, it was assumed that training on pictured representation served to: (a) strengthen generalized pencil-marking response tendencies to the similar light stimuli of the motor task, (b) strengthen response tendencies similar to the pencil-marking response, e.g., switch-pressing responses to pictured representation stimuli and to similar stimuli. The first of these assumptions requires the mechanism of stimulus generalization, while the second makes use of response generalization.

These same assumptions may be applied to the present experiment if we make the added assumption that generalization may occur along mediated dimensions. This additional assumption provides for strengthening of a generalized response tendency to a particular light stimulus by training which establishes this response tendency to a verbal symbol for the light stimulus. Such generalization, of course, depends upon previous learning in which words such as 'red' or 'top' are associated with a particular color or position so that the words have become functionally equivalent to the visual stimuli. Similarly, by the mechanism of mediated generalization, a verbal response may strengthen the motor response which it symbolizes.

The second part of the hypothesis presented previously concerns the fact that the motor task used required the learning of discriminative habits. In such discriminative learning any training which serves to reduce the internal generalization within the final task (i.e., the tendency to confuse the stimuli and responses within that task) may be expected to result in positive transfer. In the present experiment as well as the previous one training may be considered not only to have strengthened similar response tendencies to similar stimuli, but also to have reduced the range of the generalization gradients involved. If such a reduction in the range of generalization has occurred, the type of training given may have served also to increase the discriminability of the stimuli and responses of the motor task.

There is no way of telling from these experiments whether both stimulus and response generalization are necessarily important aspects of the explanation. However, it is reasonable to suppose that reduction of internal generalization among *stimuli* would be particularly important in a task where discrimination among visual stimuli plays such a major role. With an apparatus similar to the one used here, an experiment was performed to explore the transfer effects of pre-differentiation among signalling stimuli (3). Positive transfer resulted when training consisted of presenting the actual task stimuli and requiring verbal responses completely unrelated to the actual task responses (e.g., the letters *V*, *L*, *M*, or *N* were to be given in response to the lights). Thus, in a situation providing opportunity to reduce internal generalization among final task stimuli but not among final task responses, considerable positive transfer occurred. This would suggest that at least part of the explanation of the positive transfer effects of verbal training lies in the opportunity for reduction of stimulus generalization gradients. It is interesting

to note that in both the stimulus pre-differentiation experiment and the present experiment, small amounts of training had no transfer value, whereas the larger amounts did benefit the learning of the final task.

Summary

This experiment investigates the transfer effects of varying amounts of verbal training upon the subsequent learning of a motor task. Previous experiments showing positive transfer from mental practice to the performance of motor tasks have not allowed for determination of the relative roles of implicit motor responses or verbalizations. By designing this experiment so as to minimize the possibility of using implicit motor responses, we were able to study the effectiveness of a primarily verbal type of mental rehearsal. The motor task used in the experiment was a discrimination learning problem in which the subject learned to press the appropriate one of four switches upon the appearance of a red or green stimulus light in either of two positions. In the verbal training the subject was asked to memorize, by an oral paired associates method, verbal expressions of the stimulus-response relationships of the final task. Specifically, the experimenter gave two words representing color and position of light, and the subject was to respond with two other words representing the position of the correct switch. During verbal training the subject did not know that a motor task was to follow. Three groups of 18 subjects received 0, 8 and 24 verbal training trials, respectively, prior to 60 trials on the motor task.

1. When time was used as the measure of learning, no significant transfer effects were found following eight verbal training trials; but 24 verbal training trials did yield a significant amount of positive transfer to the performance of the motor task.

2. Considering the total number of errors made during the first 30 trials on the motor task, we found no transfer from eight verbal training trials, but a significant amount of transfer from 24 verbal training trials.

3. The present results were interpreted in terms of the theoretical concepts of stimulus and response generalization.

(Manuscript received August 19, 1949)

References

1. Gagné, R. M., & Foster, Harriet. Transfer of training from practice on components in a motor skill. *J. exp. Psychol.*, 1949, **39**, 47–68.
2. Gagné, R. M., Baker, Katherine E., & Foster, Harriet. Transfer to a motor skill from practice on a pictured representation. *J. exp. Psychol.*, 1949, **39**, 342–355.
3. Gagné, R. M., & Baker, Katherine E. Stimulus pre-differentiation as a factor in transfer of training. (To be published)
4. Perry, H. M. The relative efficiency of actual and "imaginary" practice in five selected tasks. *Arch. Psychol.*, N. Y., 1939, No. 243.
5. Sackett, R. S. The influence of symbolic rehearsal upon the retention of a maze habit. *J. gen. Psychol.*, 1934, **10**, 376–395.
6. Sackett, R. S. The relationship between amount of symbolic rehearsal and retention of a maze habit. *J. gen. Psychol.*, 1935, **13**, 113–128.
7. Vandell, R. A., Davis, R. A., & Clugston, H. A. The function of mental practice in the acquisition of motor skills. *J. gen. Psychol.*, 1943, **29**, 243–250.
8. Warden, C. J. The relative economy of various modes of attack in the mastery of a stylus maze. *J. exp. Psychol.*, 1924, **7**, 243–275.

SHIFTING OF DISTRIBUTION OF PRACTICE IN MIRROR DRAWING[1]

BY J. C. TSAO

Hong Kong University

Introduction

In testing the deduction from Snoddy's (2) theory, that, since "primary growth comes early and is enhanced by interpolated time . . . (and) secondary growth comes later and is enhanced by reducing the interpolated time," early spaced and later massed practice should be more efficient than early massed and later spaced trials, Doré and Hilgard (1) found that the final achievement in pursuit learning of an early concentrated and later distributed practice group was significantly higher than that of the group with the reversed distribution of trials. In a reply to Doré and Hilgard, Snoddy (3) maintained that the early spaced and later massed practice group was handicapped by interference effects which occurred because the *S* "was first adapted to time-interval conditions and then shifted to continuous practice conditions *in the same practice sitting*" and further predicted that "had these authors (Doré and Hilgard) removed the interference by interpolating an over-night interval and then measured the relatively permanent efficiency immediately following the interval they would find their two groups at the same efficiency level."

The keynote of Snoddy's contention was that interferences were due to shifting from one practice condition to another in the same sitting, and that the interference effects could be removed by interpolating an overnight interval. According to Snoddy, therefore, if the changing-over were not in the same sitting, and a 24-hour interval were to be interpolated between the practice periods, the results should be differerent from those obtained by Doré and Hilgard. In order to test this deduction, the present experiment was carried out.[2]

Experiment

The experiment of the present study was on mirror drawing, the setup of which was described in the first paper of this series (4). A brief recapitulation should suffice. The apparatus used was shown in the former paper; in general it was slightly different from the ordinary star-pattern tracing sheet. The learning score was 1/12 of the actual time, in terms of sec., taken by the *S* in each trial of tracing, which corresponded to the mean time required in tracing from the centre of the apparatus to one of the peripehal circles and back to the centre.

The *S*s were 64 R.A.F. air-crew who formed four groups of 16 people each, which were designated Groups E, F, G, and H. Groups E and F practised 12 trials in one sitting, and Groups G and H in two sittings with an interval of 24 hours interpolated between the 6th and 7th trials. In Groups E and G, trials 1–6 were spaced with one-min. intervals and trials 7–12 were continuous; in Groups F and H the procedure was reversed, i.e., trials 1–6 were continuous and trials 7–12 spaced.

Results

The average scores of each group are given in Table I and they are presented also in a graphic form in Fig. 1. Since the scores are time-records and

[1] This is the seventh article in a series of studies in spaced and massed learning. The study was carried out in the Psychological Laboratory, Cambridge, England; the writer is indebted to Mr. G. C. Grindley for help and advice.

[2] When this paper was going to be sent to the editorial office, the writer read the article of Cook and Hilgard (*J. exp. Psychol.*, 1949, **39**, 169–172), which is on the same line as the present one.

TABLE I

AVERAGE SCORES OF MIRROR DRAWING

Group	Trial											
	1	2	3	4	5	6	7	8	9	10	11	12
E	16.4	10.6	7.3	5.9	5.1	4.6	4.3	4.2	4.0	4.0	3.9	3.9
F	16.9	10.3	8.8	7.9	7.3	6.5	5.5	4.7	4.3	4.0	3.6	3.2
G	16.1	9.2	6.8	5.3	4.5	4.3	4.2	3.8	3.6	3.5	3.4	3.4
H	16.1	11.5	9.5	9.0	8.3	6.7	5.8	5.0	4.5	4.1	3.9	3.5

consequently denote the reciprocals of the speed of tracing, a larger figure in the table or a higher point on the curves means a lower efficiency.

From the scores and the learning curves, it can be seen that at the completion of the 6th trials the groups having early spaced practice, i.e., E and G, were superior to those having early massed practice, F and H. The differences between Groups E and F and between G and H in the scores of the 6th trials are significant statistically, as shown in the following:

6th trials of Groups E and F: Diff. = 6.5–4.6 = 1.9; C.R. = 4.60.

6th trials of Groups G and H: Diff. = 6.7–4.3 = 2.4; C.R. = 4.20.

With regard to final achievements, groups which had early massed and later spaced practice were slightly superior to groups which had early spaced and later massed practice. The final score of Group F was a little higher than that of Group E, and those of Groups G and H were at about the same level.

12th trials of Groups E and F: Diff. = 3.9–3.2 = 0.7; C.R. = 1.90.

12th trials of Groups G and H: Diff. = 3.5–3.4 = 0.1; C.R. = 0.44.

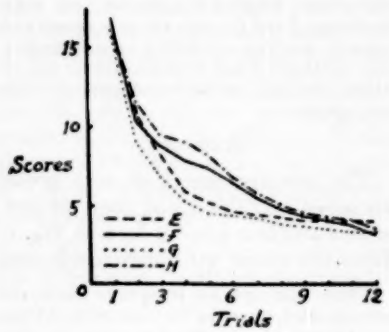


FIG. 1. Learning curves of mirror drawing

The gain in efficiency or drop in score from the 6th trial to the 7th trial was in Group E 0.3, in Group F 1.0, in Group G 0.1, and in Group H 0.9. It should be borne in mind that the performance of the 7th trial after the 6th in Groups E and G involved a shifting from spaced to massed practice, and in Groups F and H a shifting from massed to spaced practice, and that the shifting in Groups E and F took place in the same sitting, while in Groups G and H an over-night interval was interpolated. The results seem, therefore, to indicate that larger gains were made when the shifting was from massed to spaced than when from spaced to massed practice (1.0 and 0.9 vs. 0.3 and 0.1); and that it made little difference whether the shifting took place in the same sitting or after an overnight interval (0.3 and 1.0 vs. 0.1 and 0.9).

DISCUSSION

The results of Groups E and F in the present experiment agreed with

those obtained by Doré and Hilgard, which showed that early spaced and later massed practice was not so efficient as early massed and later spaced practice. The shifting from one to another form of distribution of practice in Groups G and H did not take place in the same sitting, and though the final achievements of these two groups were at about the same level the results were not fundamentally different from those of Groups E and F in which the shifting did occur in the same sitting.

When Snoddy made the statement that had Doré and Hilgard interpolated an over-night interval and measured the relatively permanent efficiency immediately after the interval "they would find their two groups at the same efficiency level," he might mean that if Doré and Hilgard measured again the efficiency on the next day after the experiment they would find that the early spaced and later massed group had caught up with the other one. It was a pity that in the present study we could not test Groups E and F in this way because of some difficulty in arranging the programme. There are some reasons, however, to doubt the effects which Snoddy attributed to an over-night interval. Firstly, the results of the present study showed that 24-hour intervals in Groups G and H did not cause greater gains from the 6th to 7th trials than the corresponding improvements in Groups E and F which did not have such long pauses. Secondly, Snoddy cited the results of an early experiment of his (2) as the evidence for the interference effects due to transference from one practice condition to another in the same sitting. In his experiment, two over-night intervals were interpolated, after the second of which there were losses in efficiency in both of the paired repetitions; and after the first 24-hour interval there was a loss in the first trial and not in the second, though Snoddy thought that there should be a loss in the second trial as well. These results may indicate that it is unlikely that the early spaced and later massed group could catch up with the rival group should they be tested the second day in the experiment of Doré and Hilgard.

According to Snoddy's theory, early massing is extremely unfavourable, because on the one hand it does not provide enough time intervals for primary growth to develop and on the other hand it may facilitate secondary growth which cannot, however, be enhanced in an early period and without an appropriate base, i.e., a certain amount of primary growth. Late spacing is also disadvantageous since the unstable secondary growth may fade away during the later intervals. Yet Doré and Hilgard found that early massed and later spaced was more efficient than early spaced and later massed practice, and the results of Groups E and F in the present study rather confirm their findings. When the conditions proposed by Snoddy were fulfilled in Groups G and H of the present study, the early spaced and later massed group still failed to show its superiority over the early massed and later spaced group. Snoddy attributed the lower efficiency of early spaced and later massed group in the experiment of Doré and Hilgard to interference effects and took the results of one of his early experiments as an evidence. In his early study, however, Snoddy interpreted the interference effects as due to the dominance of time intervals interpolated before paired trials. It is very difficult to apply the same interpretation to the results of Doré and Hilgard since the conditions in the two studies were quite different. Furthermore, time intervals in the early spaced practice could not dominate during the whole course of the later massed practice; if it did so, secondary growth could never take place.

Since, in his reply to Doré and Hilgard, Snoddy stressed the effects caused by the fact that the *S* was first adapted to time-interval and then shifted to continuous practice condition in the same practice sitting, it seems that he might mean that the interferences were due at least partly to the disturbance of the *S*'s

adaptation to the practice conditions. This is a psychological interpretation and is not congruent with his theory in general, according to which "earliness enhances primary growth, and lateness and repetition alone, the secondary form." Moreover, the results of Groups G and H of the present study made this interpretation, were it Snoddy's, unlikely to be tenable.

Summary and Conclusions

Four Groups, E, F, G, and H, of 16 *Ss* each practised on mirror drawing for 12 trials. In Groups E and G, trials 1–6 were spaced with one-min. intervals and trials 7–12 were continuous; in Groups F and H, the procedure was reversed. Groups E and F performed the 12 trials in one sitting, and Groups G and H each had an overnight interval interpolated between the 6th and 7th trials.

In final achievements, Groups G and H were about the same, and Group F slightly exceeded Group E. The results of Groups E and F agreed with those of Doré and Hilgard, and the results of Groups G and H failed to support Snoddy's supposition. The results naturally lead us to the conclusions that early massed and later spaced practice may be as efficient as, or even more efficient than, early spaced and later massed practice in mirror drawing, and that Snoddy's hypothesis regarding the interferences between primary and secondary growths is questionable. The latter conclusion may throw some doubts on the whole system of Snoddy's theory. It may be true that time intervals and practice enhance different aspects or processes of learning, which may be designated as primary and secondary growths or by any other names, but it is doubtful whether these different aspects or processes in learning follow the "rule of earliness and lateness" and are "distinctly opposite in character" and "always related to each other in terms of plus and minus," as assumed by Snoddy.

(Manuscript received
September 15, 1949)

References

1. Doré, L. R., & Hilgard, E. R. Spaced practice as a test of Snoddy's two processes of mental growth. *J. exp. Psychol.*, 1938, **23**, 359–374.
2. Snoddy, G. S. *Evidence for two opposed processes in mental growth.* Lancaster: The Science Press, 1935.
3. Snoddy, G. S. A reply to Doré and Hilgard. *J. exp. Psychol.*, 1938, **23**, 375–383.
4. Tsao, J. C. Studies in spaced and massed learning: I. Time period and amount of practice. *Quart. J. exp. Psychol.*, 1948, **1**, 29–36.

THE RELATIONSHIP BETWEEN ORDER AND FREQUENCY OF OCCURRENCE OF RESTRICTED ASSOCIATIVE RESPONSES

BY W. A. BOUSFIELD AND W. D. BARCLAY[1]

University of Connecticut

The purpose of this study was to test the following hypothesis: *for a group of Ss having a generally similar cultural background, the order of occurrence of restricted associative responses from the individuals should vary positively with the frequencies of occurrence of the responses from the group.* In other words, the most commonly given responses should occur early in individual sequences and those of lower frequencies should occur later.

The term restricted associative responses was employed by Bousfield and Sedgewick (1) to denote sequences of verbal associates given by *Ss* when their instructions stipulated the relevance of the responses. Sample instructions were: to "name as many quadruped mammals as you can"; to "name as many cities in the United States as you can"; and to "name as many pleasant objects, activities and situations as you can." Analysis of the responses indicated a negatively accelerated rate of production which the writers interpreted as being a function of the number of remaining available associates. In describing this type of situation we may postulate that the *S*, in each case, draws upon a pool of available relevant associates. As habits, these associates have been subjected to various degrees and types of reinforcement, and hence differ in strength. The reinforcements of verbal responses are largely social in character. Given a group with generally similar cultural background, we would expect a degree of similarity with respect to factors reinforcing verbal habits. Certain associative responses would be more extensively reinforced than others and hence more likely to occur. As Hull (2) has indicated, probability of occurrence is an index of habit strength. There are also grounds for assuming that habit strength should determine the order in which the *S* draws upon a pool of associative responses. This assumption may be derived as a corollary of the principle that probability of occurrence is an index of habit strength. The *S* beginning a sequence of relevant responses would be most likely to give those having greatest strength. As the sequence progresses, the supply of stronger habits becomes reduced. The conditions of the experiment inhibit repetition. The *S* must then draw upon those associates of weaker strength, and here again relative strength would determine probability of occurrence. The *S* thus tends to give his strongest associates first and his weakest last.

Assuming the validity of this analysis, both frequency and order of occurrence are related indices of habit strength, and for a group of *Ss* whose verbal habits have received more or less similar reinforcement by virtue of a common culture, these indices should correlate positively. We may further assume that the requirement of cultural similarity is adequately met by using as *Ss*, native undergraduate students (mostly with middle class

[1] The writers would like to acknowledge their indebtedness to Dr. Charles E. Osgood for suggestions leading to the design of this experiment.

and lower middle class backgrounds) attending their State University.

Procedure

The procedure for obtaining data was similar to that employed by Bousfield and Sedgewick (1). Undergraduate students in psychology courses served as *Ss*. The experiment was conducted in a large class room regularly used for the meetings of the courses. Each *S* was provided with a pencil and an 8½ × 11 in. piece of plain paper. *E*, provided with a stop watch, gave the following instructions, supplementing them whenever it appeared necessary by explanatory comments.

1. You are asked not to write your name. Please indicate your age and sex, however, on the paper that has been given you.

2. This experiment is being undertaken in order to find out something about a certain type of thinking. It will involve a problem similar to the following: write all the first names you can think of beginning with the letter *A*. Thus you might give Arthur, Alfred, Alice, Abraham, Andrew and so on. You are not being asked to work on this problem, but if you were, you would proceed by writing the names in a column at the left of the paper. At intervals I would say, "*Draw a line.*" On hearing this signal, you would draw a short horizontal line below the last name you had written, and then continue with more names. [*E* illustrates the procedure on the blackboard.] In the event you had thought of no new names since the last instruction, "*Draw a line,*" you would draw another line just the same. [*E* again illustrates the procedure on the blackboard.] Other possible instructions might be to name animals, trees, or as many parts of an automobile as possible. You are to begin writing as soon as you hear the word "Go!" Are there any questions?

Having answered such questions as appeared relevant, *E* continued with specific instructions for the production of associates. A total of 18 min. was allowed for the writing, the signals for drawing demarcation lines being given at intervals of two min. The three types of instruction were as follows:

a. For Group I (100 *Ss*): List the names of as many birds as you can. Ready, go!

b. For Group II (58 *Ss*): List the names of as many carpenter's tools as you can. Ready, go!

c. For Group III (60 *Ss*): List the names of as many celestial (that is, heavenly) bodies as you can. Ready, go!

Results

The first step in the treatment of data was to obtain the frequencies of the various associates. Table I presents the 50 most frequently occurring associates for each group. In making the frequency counts no distinction was made between singular and plural forms of associates, i.e., the lexical unit was employed. Misspellings were ignored. Inspection of the lists revealed a considerable proportion of what might be termed mediately relevant associates. These were most pronounced in the low frequency categories. In the lists of birds, there were, for example, 4 katydids, 3 duckbill platypuses, 2 gazelles, and 2 jail birds. The associates of lower frequency in Table I indicate considerable flexibility in classification.

The next step in the treatment of data was to transcribe the frequencies onto the original individual lists. The successive associates of these lists were numbered *1*, *2*, *3*, *4*, *5*, . . . *n* to indicate order of occurrence. These numbers are here referred to as the ranks of the associates. The frequencies and ranks of all associates for a group of *Ss* were then recorded on a scatter diagram using intervals of five for both frequency and rank. From this diagram the mean rank was computed for each interval of frequency. The data thus condensed appear in Table II and Fig. 1. It may be noted that there are gaps in the data for several of the frequency intervals. The data indicate, however, a pronounced tendency for associates of high frequencies to occur relatively early (to have low ranks), and for those of progressively lower frequencies to occur progressively later (to have higher ranks) in the three types of associates. The variability of the

TABLE I

50 Most Frequently Occurring Associates for Each Group

Group I (100 *Ss*) Birds		Group II (58 *Ss*) Carpenter's Tools		Group III (60 *Ss*) Celestial Bodies	
Associate	Freq.	Associate	Freq.	Associate	Freq.
Robin	98	Hammer	58	Sun	59
Crow	92	Screwdriver	56	Mars	58
Blue Bird	91	Saw	53	Jupiter	57
Eagle	90	Chisel	46	Moon	57
Sparrow	87	Nails	46	Venus	54
Canary	86	Rule(er)	45	Saturn	52
Humming Bird	85	Plane	43	Pluto	40
Blue Jay	84	Pencil	41	Big Dipper	39
Black Bird	81	Drill	40	Earth	39
Parrot	80	Screws	37	North Star	39
Hawk	79	Pliers	30	Little Dipper	33
Pheasant	75	Sandpaper	30	Milky Way	33
Pigeon	72	File	28	Mercury	29
Woodpecker	71	Level	28	Stars	22
Owl	67	Ladder	22	Neptune	21
Oriole	66	Ax(e)	20	Halley's Comet	18
Duck	60	Vise	19	Comet	15
Sea Gull	58	Wrench	18	Orion	13
Starling	55	Bolts	17	Southern Cross	13
Swallow	55	Putty Knife	15	Juno	11
Chicken	51	Knife	15	Capricorn	10
Dove	49	Saw Horse	13	Sagittarius	10
Goose	49	Chalk	13	Uranus	10
Ostrich	49	Nuts	13	Leo	9
Wren	48	Paint	13	Meteor	9
Chickadee	47	Compass	12	Planets	8
Turkey	45	Jig Saw	12	Taurus	8
Thrush	44	Paint Brush	12	Cancer	7
Peacock	44	Square	12	Morning Star	7
Kingfisher	42	Tacks	12	Aries	6
Penguin	41	Wood	12	Big Bear	6
Albatross	40	Crow Bar	10	Cassipoeia	5
Cat Bird	40	Glue	10	Clouds	5
Quail	40	Hack Saw	9	Pisces	5
Swan	40	Horses	9	Scorpio	5
Vulture	40	Mallet	9	Zenith	5
Pelican	39	String	9	Apollo	4
Partridge	38	Wedge	9	Aurora Borealis	4
Cardinal	36	Bit	8	Evening Star	4
Flamingo	36	Clamps	8	Gemini	4
Stork	36	Hatchet	8	Little Bear	4
Whippoorwill	34	T-Square	8	Northern Cross	4
Crane	33	Blue Prints	7	Northern Lights	4
Mocking Bird	33	Nail Set	7	Pleiades	4
Nightingale	33	Miter Box	7	Virgo	4
Red Wing Black Bird	33	Plumb Line	7	Alpha Centauri	3
Raven	29	Protractor	7	Betelgues	3
Bob White	28	Rope	7	Castor	3
Heron	28	Band Saw	6	Eastern Star	3
Grouse	27	Brush	6	Pollux	3

data would appear to be due in part to the sizes of the samples. For the lists of birds, 100 *Ss* had a mean of 35.97 associates per *S*. The 58 *Ss* listing carpenter's tools averaged 24.60 associates, and the 60 *Ss* listing celestial bodies averaged 17.32. Obviously other factors contributed to variability and further study would be necessary to isolate them.

TABLE II

FREQUENCY AND ORDER OF OCCURRENCE (RANK) OF ASSOCIATES IN SEQUENCES OF ASSOCIATES

Frequency of Occurrence	Group I (100 Ss) Birds		Group II (58 Ss) Carpenter's Tools		Group III (60 Ss) Celestial Bodies	
	No.	Mean Rank	No.	Mean Rank	No.	Mean Rank
1–5	310	29.0	396	19.2	247	15.6
6–10	154	25.5	197	18.5	94	12.7
11–15	122	25.8	152	15.9	51	12.2
16–20	92	26.2	75	14.2	20	14.5
21–25	129	23.7	22	17.8	42	8.7
26–30	191	24.4	118	13.4	29	10.2
31–35	168	22.9	—	—	66	11.1
36–40	389	24.7	77	11.4	157	8.8
41–45	215	23.2	129	10.8	—	—
46–50	239	22.0	94	7.5	—	—
51–55	162	19.7	53	3.9	105	6.8
56–60	118	21.0	114	6.3	228	5.4
61–65	—	—				
66–70	134	16.5				
71–75	218	18.3				
76–80	158	17.2				
81–85	248	13.5				
86–90	267	12.2				
91–95	184	9.6				
96–100	99	5.4				
Total Responses	3597		1427		1039	

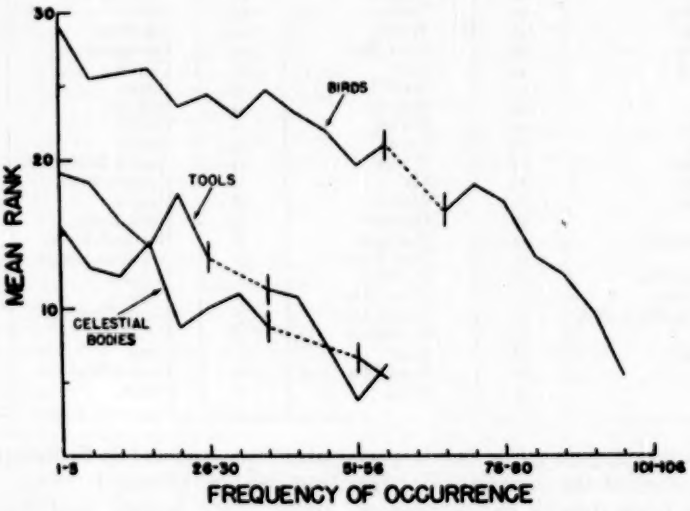


FIG. 1. Relationship between frequency and order of occurrence (rank) of associates. Gaps in the data are indicated by broken lines.

Conclusions

Assuming the validity of the experimental design and the treatment of data, we reach at least two conclusions:

1. The data support the experimental hypothesis which was formulated as follows: *for a group of Ss having a generally similar cultural background, the order of occurrence of restricted associative responses from the individuals should vary positively with the frequencies of occurrence of the responses from the group.*

2. The data indicate the feasibility of the analysis of sequences of restricted associative responses from the standpoint of learning theory.

(Manuscript received August 30, 1949)

References

1. Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

RETENTION OF PURSUIT ROTOR SKILL AFTER ONE YEAR

BY HUGH M. BELL

Department of Psychology, Chico State College

The retention of motor skill has been studied by numerous experimenters since the early work of Book and Swift on typewriting in which a high degree of retention was found even after a lapse of two years. A summary of significant studies on long-time retention in motor skills has been made by McGeoch (2).

A study (1) of the effect of rest pauses of varying lengths interpolated early and late in learning on the acquisition of pursuit rotor skill has been reported by the writer. In this investigation 457 subjects were given 20 trials of one min. each separated by a one-min. rest except where longer rests were introduced. For the early rest groups, rest periods of 10 min., one hour, six hours, 24 hours, and 30 hours were used after the fifth trial, and for the late rest groups, rest periods of the same duration were employed following the fifteenth trial. The control group used in this study rested one min. after each trial.

The present experiment consists of a retest after one year of 47 of the subjects who participated in the first experiment. Each subject was given 20 one-min. trials on the pursuit rotor separated by rest periods of one min. each. It was not possible to obtain enough subjects from any one of the

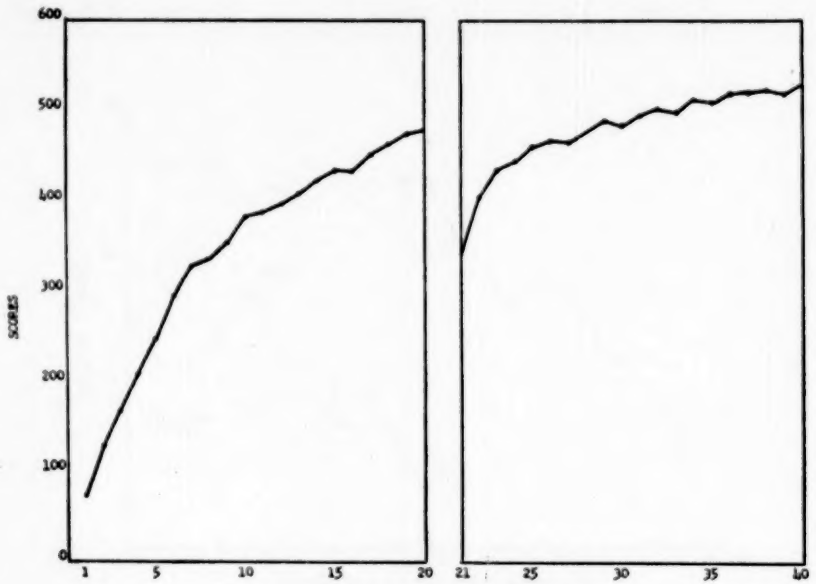


FIG. 1. Retention of pursuit rotor skill after one year. Left: Trials 1–20, April 1940. Right: Trials 21–40, April 1941

original groups and hence the scores in this experiment consist of some individuals from each of the groups.

Results

In Fig. 1 the initial learning curve of the 47 subjects for trials 1 through 20 is shown. This curve is very similar to the curve of the control group from the original experiment in that it shows an initial rapid rise followed by a gradual leveling-off as the final trial is approached.

Shown also in Fig. 1 is the retention curve of these same subjects after one year represented by trials 21 through 40. At the 21st trial (the first after the intervening year) the average score dropped to the level of trial 9 of the original series, but the next three trials show a very rapid recovery of the skill so that by the 28th trial the group has regained all that was lost over the one-year interval. From the 29th through the 40th trials the group continued to show a slight improvement in their performance on the rotor.

Analysis of the scores of these subjects indicates that they were somewhat superior to the control group used in the first experiment as their mean score at trial one was 71.9 compared with an average of 51.4 for the control group. This superiority was maintained throughout the 20 trials of the original experiment and at trial 20 their average was 479.7, 23 points above that of the control group.

After the year interval, the initial trial of the present experiment, trial 21, shows a drop to 340.4, a loss of 139.3, or 29 percent. By trial 28 the mean was 479.4, the point reached by the group at the final trial of the original learning series. From trials 29 through 40 the group gained an additional 49 points, reaching an average score of 529.9 at the final trial.

Conclusions

Over the one-year period following 20 one-min. trials scores dropped only 29 percent, and this was completely recovered after eight trials. Slight gains continued during the remaining trials. The long-lasting results of a brief practice period are convincingly demonstrated for this motor skill.

(Manuscript received October 3, 1949)

References

1. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, 54, No. 1.
2. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, and Co., 1942, pp. 328–330.

SOME EFFECTS OF SHAPE ON APPARENT BRIGHTNESS *

BY R. M. HANES

The Johns Hopkins University, Institute for Cooperative Research

Introduction

While there is considerable contradictory evidence, numerous studies have produced results which indicate that the threshold for recognition of a visual test object is influenced to some extent by the shape of the object. Kleitman and Blier (5), for example, found that triangles, circles, squares, and stars of approximately the same area can be discriminated in the order given. For binocular discrimination of two objects presented simultaneously Whitmer (7) found that the rank of different shapes based on the percent of correct discriminations was as follows: triangle best, then diamond, square, rectangle, circle and hexagon. King, Landis, and Zubin (4), using only squares, triangles, and circles, found subliminal perception (forced guesses) to exceed chance. While the results for the various figures did not differ significantly, the triangle did show up as somewhat better than the other two. Believing that the conflicting reports of previous investigators were due to lack of consideration for area of figure and specific framework effects, Hochberg, Gleitman, and Macbride (3) attempted to maintain better control over these factors. In this case threshold for form was found to increase with decreasing simplicity or compactness, as defined by the ratio of perimeter to area. The circle was found to have the lowest threshold, the square next, and the cross the highest. Langstroth *et al.* (6), on the other hand, report that there has been no difference found in the probability of detection for square, rectangular, triangular, and disc-shaped targets when target area is constant.

While this brief review is not exhaustive, it serves to indicate that limited agreement has been found. Triangles, perhaps, have been most persistent in giving lower thresholds. Whether threshold results, even if consistent, should prevail above threshold is open to question; the threshold experiments are cited primarily as being suggestive. But if 'compactness' is a factor in detection at threshold, it might also help to determine apparent brightness. If apparent size is a determiner of threshold for form, it might also be a factor in apparent brightness. The present study was undertaken in an attempt to answer some such possibilities.

Apparatus

Two luminous opal glass screens, presented simultaneously, served as the stimuli. The light source for each was a 150-watt projection bulb housed in a metal box. The light from each bulb passed successively through a first sheet of opal glass, a diamond-shaped diaphragm, and a second sheet of opal glass, this latter glass being the one viewed by the observer. The different stimulus shapes and sizes were obtained by mounting photographic slides on the opal glass. Each shape and size was reduced photographically from an enlarged (ten to one) original. Circles, triangles, and squares were the forms used. Each was made in three different areas: 0.0031, 0.0123, and 0.7854 square in. These three areas were determined from circles of $\frac{1}{16}$, $\frac{1}{8}$, and one in. diameters, which at the viewing distance of 24 in. subtended visual angles of about 9, 18, and 144 min., respectively. The $\frac{1}{8}$-in. size was used

* This research was carried out under Contract N5-ori-166, Task Order I, between Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-109, Project Designation No. NR-784-001 under that contract.

to make certain that shape would be clearly recognized even at the higher brightness levels; the $\frac{1}{16}$-in. offered a slight possibility of confusion.

The two stimuli were located about 4½ in. apart at the end of a rectangular (3½ by 8 in.) box lined with black velvet, whose average reflectance was less than one percent. This box was attached to a large metal bowl at the end opposite the stimulus discs; thus, the *O* could see only the two luminous stimuli surrounded by a dark field. Brightness calibration was made with the use of a Macbeth Illuminometer.

The brightness of each stimulus could be varied continuously either by *O* in front of the apparatus or by *E* at the rear. Wratten neutral density filters were used for gross changes in intensity level.

Procedure

O was seated comfortably in front of the apparatus upon entering the dark experimental room and given several practice trials to acquaint him with the nature of the task. He was then presented with two stimuli of either the same or different shape but of the same area and told to adjust whichever one served as variable until its brightness was equal to that of the standard. No time limit was imposed on the trials. Each shape was compared with itself and both other shapes of the same area, as both standard and variable, on both the right and left, at each of three brightness levels: 0.1, 10, and 100 mL. There were thus 162 comparisons for each *O*. The stimuli were extinguished between comparisons while a new pair was being set.

Observers

Five male college students, between 20 and 25 years of age, served as *O*s. All had normal visual acuity or were corrected for defects. None had any experience in psychological experiments.

Results

The principal data are presented in Table I, where the results for five observers have been averaged. The table is to be read as follows: When the standard was a triangle having an area of 0.0031 sq. in. and a luminance of 0.1 mL., the mean value for an equal-sized square as variable which appeared equally bright was 0.125 mL., but for a triangle the mean value was only 0.095 mL. The triangle, therefore, appeared brighter than the square. The means of the rows, along the right side of the table, show the average values required for each indicated shape to match a composite standard at each indicated luminance. The column means in each section of the table represent the average values required to match the indicated standards. The row means reveal that, when an average is taken over all sizes, the square always appears least bright and the triangle and circle generally do not differ very much. The column means indicate, however, that while the same relationship holds true in the majority of cases, the shape which appears brightest is somewhat dependent upon the size and brightness level.

The settings for each observer under each of the various conditions were expressed as percentages of the standard and analyzed by means of the analysis of variance. This analysis revealed that the over-all differences mentioned above among the three shapes were significant beyond the 0.1 percent level of confidence. In other words, the square was significantly less bright than the triangle and the circle. A *t*-test showed that the difference between the apparent brightness of the triangle and the circle was not significant. The differences due to brightness level were also found to be highly significant; again a *t*-test showed that this effect was due primarily to the difference between the 0.1 mL. level and the other two. This same type of change in constant error in this kind of test situation, has been noted before (1, 2). The over-all differences among *O*s were not significant, and the main effects due to the different sizes were barely significant (about the five percent level of confidence).

The first-order interactions between *O*s and both size and brightness were

TABLE I

RELATIONSHIP BETWEEN APPARENT BRIGHTNESS AND SHAPE FOR THREE SIZES OF STIMULI AT THREE BRIGHTNESS LEVELS

Each entry in the body of the table is the mean value in millilamberts for five *Os* which appeared equal in brightness to the indicated standard.

	Standard									
Shape	Square			Triangle			Circle			
Area (sq. in.)	.0031	.0123	.7854	.0031	.0123	.7854	.0031	.0123	.7854	
Variable	0.1 mL. level									Mean
Square	.110	.106	.101	.121	.125	.117	.129	.099	.132	.116
Triangle	.099	.096	.101	.095	.100	.104	.105	.091	.124	.102
Circle	.104	.104	.093	.108	.123	.102	.091	.103	.105	.104
Mean	.104	.102	.098	.108	.116	.108	.108	.098	.120	
	10 mL. level									
Square	9.5	10.2	10.2	10.6	11.0	7.9	9.6	10.2	13.1	10.3
Triangle	9.2	9.2	10.1	9.5	9.8	9.1	9.2	8.7	12.8	9.7
Circle	9.3	9.8	6.8	12.4	10.1	7.0	9.9	10.3	9.2	9.4
Mean	9.3	9.7	9.0	10.8	10.3	8.0	9.2	9.7	11.7	
	100 mL. level									
Square	113	101	93	102	111	98	113	108	135	108
Triangle	114	91	94	124	88	101	94	81	115	100
Circle	87	95	66	115	102	63	105	101	94	92
Mean	105	96	84	114	100	87	104	97	115	

significant, however. In other words, different *Os* varied in their matches depending upon the particular size and brightness involved. On the other hand, the interaction between *Os* and shape was only barely significant at the five percent level of confidence. The interaction between size and shape was, however, highly significant, indicating that the noted differences among the column means in Table I, as a function of size, are not chance variations. In fact, the sums of squares for the interaction of size and shape were larger than any of the other sums. Which shape appears brightest, therefore, is largely dependent upon the size involved in the comparison. Nevertheless, as indicated by the column means in Table I, only in one of the comparisons does the square show up as brighter than the triangle. The superiority in apparent brightness for the circle as compared with the square is more equivocal.

Table II is presented for the purpose of allowing for a more readily discernible picture of the effects of shape on apparent brightness. In this table each entry is the mean of the three size means for each standard and variable at each brightness level. The greater apparent brightness of

TABLE II

RELATIONSHIP BETWEEN APPARENT BRIGHTNESS AND SHAPE AT THREE BRIGHTNESS LEVELS

Each entry in the body of the table is the mean value in millilamberts of five *Os* which appeared equal in brightness to the indicated standard.

Variable	Standard: Square	Triangle	Circle	Mean
0.1 mL. level				
Square	.106	.121	.120	.116
Triangle	.099	.100	.107	.102
Circle	.100	.111	.100	.104
Mean	.102	.111	.109	
10 mL. level				
Square	10.0	9.8	11.0	10.3
Triangle	9.5	9.5	10.2	9.7
Circle	8.6	9.8	9.8	9.4
Mean	9.4	9.7	10.3	
100 mL. level				
Square	102	104	119	108
Triangle	100	104	97	100
Circle	83	93	100	92
Mean	95	100	105	

the triangle and circle as compared with the square is more clearly shown here by both the row and column means. There is no significant interaction between rows and columns, thus indicating that the same relationships for shape and apparent brightness are found regardless of which shape is used as the standard.

DISCUSSION

It seems quite clear that the form of a luminous stimulus can influence the apparent brightness of that stimulus. And while no one shape (of the three studied here) has been found to appear brighter than the others under all conditions tested, nevertheless the triangle, and probably also the circle, definitely appears brighter on the average than does the square. If any superiority in threshold sensitivity exists for form, it is probably the triangle which has been shown most often to give the lowest thresholds. On this point, at least, there would seem to be some agreement between threshold and supra-threshold effects.

Why should these differences in apparent brightness occur? Since area remains the same for the three shapes, explanation for the effects must be based on differences in perimeters, maximum dimensions or apparent size. It seems that all three possibilities must be invoked to account for the main effects and interactions found here. The enhanced brightness of the triangle at the two smaller sizes could be accounted for by the longer dimensions of the triangle relative to the square and circle. If border contrast were a primary factor in determining apparent brightness, and het total extent of a given contrast also influenced apparent brightness, then the triangle should appear brightest. That the triangle does not also appear brightest when the largest size is considered would make this hypothesis questionable except that a further factor may become operative at this size. It was determined in a previous experiment that apparent brightness increased with increasing size at low intensity levels (0.1 mL.) but that this relationship was in general reversed at higher levels (100 mL.). Now with the two smaller sizes studied in the present experiment there is a negligible difference in apparent size among the three shapes. When the largest size is considered, however, the triangle looks considerably larger than the other two. Therefore, with the large size and higher intensity levels,

the triangle, since it looks larger, might well have its apparent brightness diminished. The data substantiate, to a degree at least, such a hypothesis. The fact that the corners of the triangles would extend farther into the periphery than would the edges of either the square or circle, and would therefore stimulate more rods and less cones, could also cause the apparent brightness of the triangle to be reduced relative to the other figures at higher intensity levels.

These explanations do not satisfactorily allow for the marked difference between the circle and the other two at the 0.1 mL. level and largest size. No satisfactory explanation for this discrepancy seems apparent. It is obvious that these various relationships are quite complex and that no single explanation is adequate for the influence of form on either thresholds or apparent brightness.

Summary

The influence of shape or form on apparent brightness has been investigated for three geometrical forms (square, triangle, and circle) at three intensity levels (0.1, 10, and 100 mL.). Comparisons were made only among forms of equal area.

The triangle and circle, on the average, exhibited an enhanced brightness relative to the square. These differences do not, however, remain the same for all sizes and all brightness levels. In other words, there are some significant interactions among the three factors. When the two smaller sizes are considered the triangle has a consistently higher apparent brightness than do the other two. For the largest size the circle appears to be brightest.

While the results can be partially explained in terms of marginal contrast and apparent size, no simple single explanation seems adequate to account for the results.

(Manuscript received for immediate publication July 7, 1950)

References

1. Hanes, R. M. The construction of subjective brightness scales from fractionation data: a validation. *J. exp. Psychol.*, 1949, **39**, 719–728.
2. Hanes, R. M. Supra-threshold area-brightness relationships. (In press)
3. Hochberg, J. E., Gleitman, H., & Macbride, P. D. Visual threshold as a function of simplicity of form. *Amer. Psychologist*, 1948, **3**, 341–342. (Abstr.)
4. King, H. E., Landis, C., & Zubin, J. Visual subliminal perception where a figure is obscured by the illumination of the ground. *J. exp. Psychol.*, 1944, **34**, 60–69.
5. Kleitman, N., & Blier, Z. A. Color and form discrimination in the periphery of the retina. *Amer. J. Physiol.*, 1928, **85**, 178–190.
6. Langstroth, G. O., *et al.* *A survey of visibility lumen data.* Dept. of Physics, U. of Manitoba, Proj. C.E. 128, Dept. of National Defense, Jan. 18, 1945. Pp. 20.
7. Whitmer, C. A. Peripheral form discrimination under dark adaptation. *J. gen. Psychol.*, 1933, **9**, 405–419.

ACCURACY OF VISUAL INTERPOLATION BETWEEN SCALE MARKERS AS A FUNCTION OF THE NUMBER ASSIGNED TO THE SCALE INTERVAL[1]

BY A. CHAPANIS AND M. LEYZOREK[2]

Department of Psychology, The Johns Hopkins University

Introduction

During the past several thousand years, mathematicians and logicians have built up a considerable body of information about the formal properties of numbers and their uses. In spite of this, very little is known about the psychological problems involved in the handling of numbers when they are attached to tools of everyday life such as scales of length or weight. That there are some genuine psychological problems here can be inferred from a few studies already published. There is now clear evidence, for example, that most people have marked number preferences and use some numbers more commonly than others. Thus, if simple numbers are attached to marks on a ruler and people are asked to interpolate the final digits of a series of measurements, the distribution of final digits does not come out as one would predict. Even when the digits from 0 to 9 have equal probability of occurrence, too many 0's, 2's, 5's and 8's are selected and the digits 1, 3, 4, 6, 7, and 9 are largely avoided (8).

Of somewhat greater interest are the results of a war-time report by Vernon (5) who was concerned with some factors contributing to ease of scale and dial reading. Among other things, she varied the number of divisions between numbered markers as well as the numbers assigned to certain of the markers. Her results show that both factors have a strong influence on the ease and precision with which the scales could be read. Scales in which the smallest unit was equal to 1, 2, or 5 were read more accurately than scales in which the smallest unit was equal to 4, 2.5, or 8. Vernon tested only a few possible combinations of these factors so that her results are not as general as we might like.

A related study by Bartlett *et al.* (1) investigated the precision of visual interpolation in a polar-coordinate plot as a function of the number assigned to the scale interval. Only three number intervals (5, 10, and 20) were used, but the results show an unmistakable difference between the accuracies attainable with the three scales.

Purpose of the Present Experiment

The experiment reported here explores somewhat more thoroughly than has been done heretofore the relationship between accuracy of visual interpolation and the number assigned to the scale in which estimations are made.

Apparatus

Four standard naval radar indicators were used as experimental equipment. These indicators were of the plan position type, now widely

[1] This investigation was performed under Contract N5-ori-166, Task Order I, between the Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-107, Project Designation No. NR-784-001, under that contract. A preliminary account of this experiment was reported by Chapanis (2) at the 1947 Meetings of the American Psychological Association.

[2] Despite the large number of machines and men involved and the complex procedures necessary, this experiment was completed smoothly and efficiently. Credit for this accomplishment is due in large measure to the diligent and cooperative personnel of the Field Laboratory and the Naval Training Facility (Air Intercept) Beavertail who assisted in the work. We deeply appreciate the help given us by both the Naval Officers and Radarmen of the Facility and the civilian technicians and engineers of the Field Laboratory.

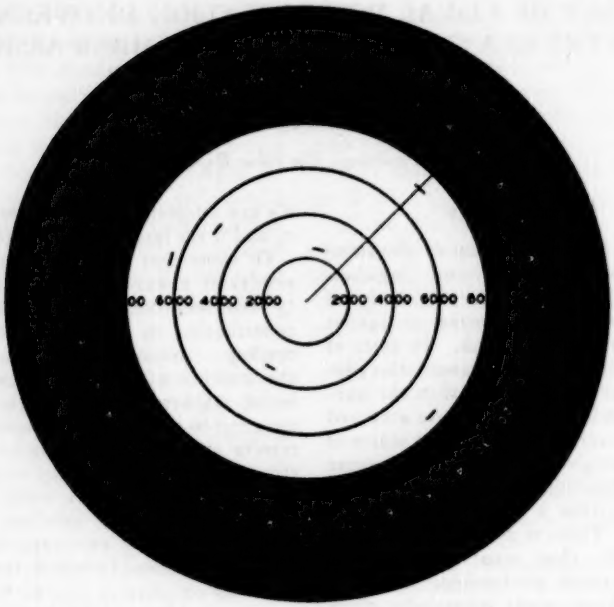


FIG. 1. A schematic diagram of the polar coordinate display showing the concentric circular scale markers, the points whose positions are to be estimated, and the circumferential angular scale. Visual interpolation is attempted along any radius. The circular scale markers are numbered here for a trial with the 2,000-yard interval. Thus, there is a point at about bearing 046° and at a distance of about 7,100 yards. The brightness relations are reversed here; actually the field is dark and the scale markers and estimation points are bright (see the text for a fuller explanation).

used as aids in navigation on the sea and in the air.[1] Although somewhat unorthodox for psychological experimentation, these machines lend themselves well to the presentation of a variety of visual problems. Since the details of the design and construction of the radar indicators are of no consequence to this report, the equipment will be described only in the most general terms, but the psychological characteristics of the problem as presented to the subject will be discussed in detail.

The Visual Display

The face of a cathode-ray tube (CRT), approximately 10 in. in diameter, was the field on which the subject's visual problem was presented. The subject saw a pattern of four thin, phosphorescent, equally-spaced concentric rings on an unsaturated, dark, gray-green field. The rings served as circular scale markers. Interspersed among the scale markers and concentric with them were several short, bright, phosphorescent arcs (see Fig. 1). These were the points to be located by visual interpolation between the appropriate scale markers.

The points and the concentric scale markers were 'painted in' on the surface of the CRT in the wake of a thin, bright, phosphorescent radial line which rotated about the center of the circular field at six rpm. This entire display was the visible evidence of a pattern of excitation imposed on the phosphor coating of the CRT face by a beam of electrons rapidly scanning the surface of the tube from the center to the periphery and back while rotating about the center. (So fast was the scanning rate of this beam of electrons that it was seen, for the most part, as a single radial line rotating about the end of the line located at the center of the CRT.)

The basic color of the pattern of concentric circles and short arcs was bright blue, but it varied somewhat in saturation with the intensity of the excitation imposed on the phosphor screen and

[1] For a more detailed and technical description of the type of equipment used in this experiment consult recent articles by Williams *et al.* (6, 7).

the time after excitation. At the sweep line, the pattern was brightest, decaying with increasing distance (and time) behind the sweep to a yellow-green and finally to the ground color of the circular field. In the interval between successive sweeps (10 sec.) the brightness of the pattern diminished slowly, but the scale markers and the points between them remained visible throughout each rotation cycle of the sweep line. With each rotation of the sweep the pattern was simply renewed in brightness.

The circular scale markers were approximately one mm. thick. Measured along any radius of the display, the points used for localization varied from approximately two to three mm. in thickness and were six degrees wide. Five points were presented in the circular field at one time. Because of electronic limitations, no points were ever presented in the innermost area of the display.

Together with its accessory equipment and necessary circuits, the CRT was mounted in an enclosure about 36 in. high and 17 in. square. The horizontal viewing surface of the tube was visible through a protective cover glass in an opening in the top of the enclosure. Between the cover glass and the face of the tube were mounted an azimuth scale laid off in degrees around the edge of the tube face (see Fig. 1).

Arrangement of the Apparatus

Two identical cathode ray tube units of the kind described above were used to enable two subjects to work at the same time. For convenience these will now be called test units. These units were linked in a circuit with two high-precision display units equipped with devices for accurately determining the true position of each localization point in the circular test field. These precision units furnished the criterion information for determining the actual position of each point—hence, these will be called criterion units. All four display units (two test, two criterion) were supplied simultaneously with the same scale and point pattern by a complex signal generator (SG) of special design. A block diagram of the circuit connecting these units is shown in Fig. 2. The four display units were isolated from each other to prevent the subjects and the criterion unit operators from hearing each other's reports on the position of any point. Because of this equipment isolation, a telephone network was used to link the personnel working at the various stations to a central point at the SG without allowing them to hear each other. The subjects were not provided with phones.

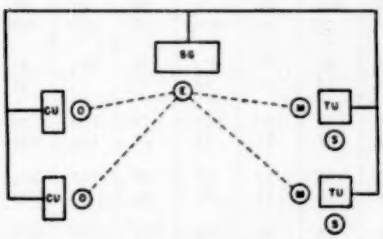


FIG. 2. A schematic diagram of the arrangement of equipment and personnel used in this study. Shown here are the test units (TU), criterion units (CU), signal generator (SG), operators (O), subjects (S), monitors (M), and experimenter (E). The solid lines show linkages between equipment; the dotted lines show telephone connections between the experimenter, operators, and monitors.

PROCEDURES

After setting in a given pattern the experimenter notified the criterion unit operators and the monitors standing by at the test units that a pattern was ready for use. The criterion unit operators immediately began their work of reporting to the experimenter the exact position of each point in the array, always identifying the point by its angular location in the field. The experimenter recorded the data thus furnished him by each criterion operator and maintained a running check on the agreement between the two equipments and their operators. In general, the agreement between the reports from the criterion instruments was very high, with discrepancies of the order of 0.05 percent. When the reports differed by more than 0.5 percent the experimenter requested a recheck from each criterion unit operator. If the discrepancy consistently exceeded 0.5 percent after checking, the experiment was halted until the equipment could be examined and recalibrated.

When the experimenter indicated that a pattern of points was ready, the monitors standing by the two test units started the subjects in their task. Each monitor specified the number system assigned to the scale markers for that particular trial, and, using a grease pencil, wrote in on the glass cover of the display the numerical value of each marker (see Fig. 1).

Eleven different numerical units were tested: 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, and 10,000 yards, and 2.5 miles. For example, in one trial the subject was told that the interval between any two scale markers represented 2,000 yards and that he was to make his estimate of the position of each point in terms of such a scale unit. In this case, the innermost marker in Fig. 1 would represent 2,000 yards,

the next 4,000 yards, the next 6,000 yards, and the outer edge 8,000 yards. Since the nautical mile is equivalent to 2,000 yards, a 2.5-mile scale unit was included to evaluate its psychological equivalence to the 5,000-yard scale. Accuracy of judgment was emphasized as the primary objective. No time limit was imposed for the task.

Starting from an arbitrary reference radius of the circular display, the subject took each point in turn, estimating the position of the point relative to the center of the field and to the scale markers between which it fell. His estimate was reported orally to the monitor who stood beside him as he worked. When both subjects (one on each of the test units) finished with a group of five points, their monitors called the experimenter to indicate readiness for the next pattern. Five such cycles (of five points each) constituted a complete trial. The average time per complete trial was about 35 min., with some variation from 25 to 45 min.

Experimental Design

The experimental trials were arranged in the form of an 11 × 11 two-factor design with one replication, so that each subject was tested twice with each of the 11 scales. The sequence of trials was arranged so that the scales were used in a different order for each subject in each series. Each series of trials (that is, the original and replication) was run on a separate test unit and the subjects alternated between test units (and thus between the trials of each series) in a random order.

TABLE I

THE EXPERIMENTAL DESIGN USED IN THIS STUDY

The entries are the trial numbers assigned to each condition. For example, on the first trial subject J was tested on instrument 1 with the 4,000-yard scale; on the second trial, subject B was tested on instrument 1 with the 8,000-yard scale and subject H was tested simultaneously on instrument 2 with the 2,000-yard scale.

Subject	Instrument	Number Scale										
		1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000	10,000	2.5
A	1	44	100	139	124	14	23	6	131	140	121	94
	2	41	28	26	9	60	105	73	97	81	46	138
B	1	116	110	21	55	142	143	72	2	127	80	50
	2	91	107	113	85	12	64	128	52	69	18	34
C	1	76	4	145	95	39	77	27	101	11	24	119
	2	50	74	103	98	32	118	79	66	57	7	42
D	1	9	13	42	98	149	106	150	74	25	47	82
	2	77	95	4	23	127	29	39	101	58	126	125
E	1	7	22	146	73	84	97	108	120	123	125	41
	2	102	100	76	65	49	44	31	27	10	6	94
F	1	79	112	17	51	87	28	126	35	36	109	49
	2	133	136	106	38	54	71	63	11	20	115	137
G	1	16	141	111	89	105	10	34	48	56	117	114
	2	3	92	53	129	62	36	33	86	19	70	84
H	1	38	32	19	83	113	52	91	69	85	15	137
	2	16	2	72	104	88	61	55	116	110	35	47
I	1	128	30	86	53	92	102	20	144	3	81	70
	2	117	114	132	89	130	17	13	56	111	67	108
J	1	31	46	130	1	115	37	103	18	71	138	90
	2	109	82	68	87	51	131	134	14	135	112	21
K	1	104	129	75	33	5	45	96	148	122	40	12
	2	59	22	25	99	43	8	147	78	93	119	80

For various reasons, such as breakdown of equipment, and loss of a subject in the middle of the experiment, we could not adhere rigidly to the experimental design and randomized sequences made up before the experiment began. Table I shows the design we actually used. It is, we feel, completely adequate for our purposes.

The experiment extended over 24 working days with 3 to 12 trials run per day. On 92 trials, two subjects were run simultaneously on the two test units as originally planned; on 58 trials only one subject was tested.

Patterns of localization points were selected randomly within prescribed areas so that the entire surface of the visual display was equally well sampled for each subject and for each scale.

Subjects

The subjects were 11 experienced naval radar operators trained in the use of polar coordinate plots and familiar with the process of visual interpolation in circular scales.

Results

The true position of each point was defined as the mean of each pair of determinations made with the two criterion units for each localization point. After appropriate conversion (in terms of the correct numerical unit for each trial) the mean criterion position for each point was subtracted from the subject's estimate. The deviation thus obtained was expressed as a percentage of the interval used in that trial. For example, let us suppose that a subject, on a trial with the 2,000-yard scale, reported the position of a point as 5,000 yards. If the true position of the point was 4,960 yards, the deviation is +40 yards. The relative error score is +40/2000 or +2.00 percent. Twenty-five such scores were obtained in each trial.[4]

Constant Errors as a Function of Distance from the Center of the Display

When the entire batch of data was in, errors were first plotted as a function of the distance of the localization point from the center of the display. The results, shown by the solid line in Fig. 3, reveal two marked trends. First, there are large positive errors near the center or innermost ring of the display and these gradually become less positive for points near the outer edge of the display. Second, there is a marked cyclical variation in the errors between each pair of scale marks.

Without elaborating, we shall state categorically that the first trend can be traced to calibration errors between the test and criterion units. Daily engineering protocols maintained by the electronic technicians show these errors consistently. Unfortunately, we, as psychologists, understood so little of the electronics of the radars that we were not able to anticipate and correct for the constant errors in the equipment calibration procedures. The straight line drawn through the data in Fig. 3 was fitted by least squares to the entire group of data and was used to correct for the constant errors. Constant errors, with the calibration errors removed, are shown by the interrupted line in Fig. 3. In making this correction, we are not exactly sure what the zero calibration error should be. The correction for overall *trend* is valid, but the point of true zero error is open to some question. In statistical terms, we are sure of the correction for slope but are not sure of the exact correction for the intercept constant. We have arbitrarily equated the mean error for the 11 different numerical scales to the intercept constant. It might be, however, that all or most of the numerical scales would show positive constant errors—if we could measure them exactly. If this were the case, the intercept constant we selected was too low. In the converse case, all or most of the

[4] In a few instances, the N for a trial is 24, due to recording errors.

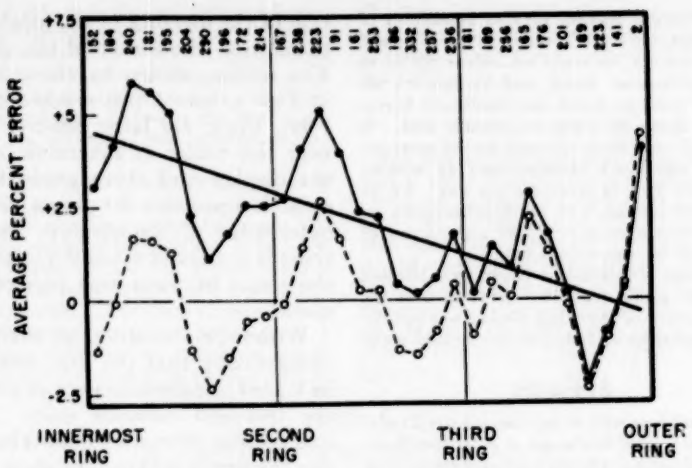


FIG. 3. The solid line shows constant errors of estimation as a function of the radial distance of points from the innermost scale marker to the edge of the display. The N at each point is shown at the top of the chart. The straight line drawn through these data is fitted by least squares and has the equation $M_e = -0.165\ R + 4.599$, where R varies from 1 (on the left) to 30 (on the right). This straight line shows the constant errors of calibration between the criterion and test units. When these constant errors are eliminated, the constant errors of estimation are shown by the dotted line.

numerical scales might have negative constant errors, in which case our correction for the intercept constant was too large. This is not as serious a difficulty as might appear at first glance, and the interested reader is referred to Chapanis *et al.* (3) for a more thorough discussion of the measurement of errors in systems such as these. In any case, the intercept selected in this experiment is in error by no more than about one percent.

The cyclical nature of the error data evident in Fig. 3 is shown more clearly in Fig. 4 which combines all the data into tenths of the interval between any two scale marks. This figure shows that points between an inner reference ring and the midpoint of the interval tend to be overestimated; points between the midpoint and the outer reference ring tend to be underestimated; points near reference rings, or near the midpoint of the interval, tend to have small constant errors. In a study of a paper display something like the one we have used, Bartlett *et al.* (1) found no clearcut variations in constant errors such as we report here. This may be due to the fact that Bartlett *et al.* have discontinuous data (subjects were required to report their estimates in units which were 1/20 of the interval between reference rings), whereas our data are continuous (subjects were required to report their estimates as precisely as possible). With the discontinuous scale used by Bartlett, the only possible errors were those of 0, 5, 10, 15, etc., percent. In our data, errors could assume any magnitude. Note too that our constant errors, though consistent, are small in amount —they do not exceed two percent. It is possible, therefore, that this tendency did not show up in the Bartlett

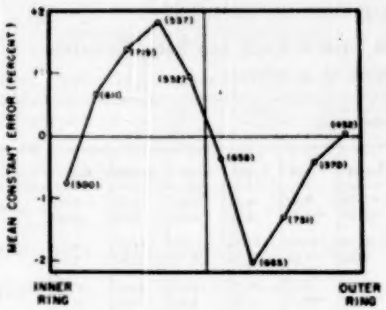


FIG. 4. The data in Fig. 3 (interrupted line) are here combined for the three scale intervals. The numbers in parentheses are the numbers of observations contributing to each point. This figure shows a consistent pattern of over- and underestimation according to the location of the point. Note that points near the midpoint of the interval are estimated with virtually no constant error.

study because of the coarseness of the grouping forced on the judgments of the subjects.

Variable Errors of Interpolation as a Function of Location Between Reference Markers

There is also a marked variation in the variable errors as a function of the position of the point between reference rings. In Fig. 5, we have plotted the standard deviations of the relative errors of estimation for all numerical scales as a function of the position of the points between reference rings. Subjects showed much less variability in their estimates for points near reference rings. They also showed less variability in estimating points near the middle of the interval. Variable errors for points in the outer half of the interval are greater than for those in the inner half. Bartlett *et al.* (1) showed similar data for variable errors in their experiment and the agreement between their curves and ours is good. They found, as we did, that variable errors are smallest near reference rings and in the center. They also found variable errors in the outer half to be greater than those in the inner half and, finally, their standard deviations (when multiplied by five to convert from their units of measurement into our 'percent errors') are of about the same magnitude as ours. In summary, then, the data in Figs. 4 and 5 both show that the midpoint of the interval has some sort of psychological stability and reference value.

Constant Errors for the 11 Number Scales

The primary purpose of this study was to investigate the effect of different number scales on accuracy of interpolation. Accuracy here has been analyzed into two parts: constant errors and variable errors. Table II shows the mean constant errors for the 25 observations made by each subject under each condition of experimentation. The analysis of variance for these data is shown in Table III.

It is clear that there are marked differences between the constant errors

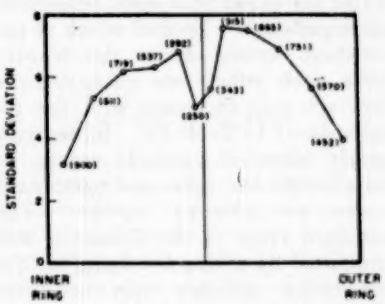


FIG. 5. Standard deviations of relative errors of estimation as a function of the radial distance of points between scale markers. The numbers in parentheses are the numbers of observations contributing to each point. The precision of estimates is high near scale markers and near the mid-point between markers; at other places it is not as good.

TABLE II

MEAN RELATIVE ERRORS OF ESTIMATION FOR EACH SUBJECT AND EACH CONDITION

The entry in each cell is based on 25 estimates.

Subject	Instrument	Number Scale											Mean
		1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000	10,000	2.5	
A	1	1.21	−0.34	−4.66	0.53	1.45	−1.14	−0.95	0.38	1.97	0.51	−0.07	−0.06
	2	1.55	−1.64	−0.74	−2.01	3.37	−0.70	−3.32	0.69	−0.05	2.62	0.06	
B	1	−0.25	−0.45	−0.40	5.45	−0.32	0.70	−0.92	−0.35	2.37	−0.01	−0.70	0.48
	2	0.62	0.20	−1.89	1.28	−1.86	2.34	−0.90	5.52	2.68	−0.86	−1.67	
C	1	−0.78	−1.46	2.01	−0.59	−0.53	0.52	1.48	−0.26	1.71	0.06	0.95	0.32
	2	3.32	0.64	0.17	0.85	−1.83	−1.01	−0.28	1.47	3.45	−2.95	−0.82	
D	1	2.15	0.64	−0.68	1.16	0.34	−1.48	0.78	−2.02	−0.99	0.40	−3.70	−0.04
	2	0.43	−0.10	1.31	−0.93	0.78	−0.99	2.76	0.52	1.34	−0.44	−2.17	
E	1	−3.97	1.16	2.08	0.57	−0.80	−0.08	−3.85	−0.10	1.37	3.07	−3.92	0.36
	2	1.01	2.26	−1.77	4.30	2.63	2.89	2.78	3.10	−0.85	0.81	−4.79	
F	1	0.12	−1.94	−13.99	2.82	−0.88	1.66	−0.81	1.85	−0.09	1.35	−2.31	−1.05
	2	−0.08	−3.48	−5.14	1.39	0.51	2.92	2.06	0.18	2.04	−0.79	−10.48	
G	1	0.56	0.02	0.12	−0.74	0.26	3.15	0.12	1.76	4.08	−0.19	−0.99	0.46
	2	−1.27	0.66	1.90	−0.55	0.51	3.76	0.27	−0.86	1.68	3.13	−7.31	
H	1	0.28	0.44	−2.27	2.11	−1.44	7.01	−0.02	1.92	−1.90	−2.96	2.88	0.37
	2	−1.12	0.02	2.15	−0.82	−1.14	−0.42	2.21	−1.74	0.31	1.66	1.04	
I	1	0.98	−0.19	−0.88	3.39	−0.98	−0.60	0.68	0.43	0.35	0.72	−3.34	−0.04
	2	−0.07	−0.30	−0.40	0.44	−0.10	−1.43	−3.97	3.47	−0.16	3.75	−2.82	
J	1	−1.45	0.34	−0.12	1.57	−1.56	0.98	−1.52	0.29	0.64	−2.84	−6.46	−1.00
	2	−0.68	−2.04	5.89	−0.53	2.48	−2.17	−3.43	0.62	−1.30	−1.43	−9.38	
K	1	−1.55	−2.06	0.57	−0.62	0.73	−0.81	−1.75	−0.36	1.05	−0.67	0.09	−0.59
	2	1.30	−1.46	−0.67	0.52	1.06	−1.80	−2.83	−0.22	−0.87	−2.04	−0.40	
Mean		0.10	−0.41	−0.79	0.89	0.12	0.60	−0.52	0.74	0.86	0.17	−2.56	−0.07

Mean (Instrument 1) = −0.11; Mean (Instrument 2) = −0.04.

for the different numerical scales, and Table III shows that these differences are significant. To find which of the constant errors differ significantly from each other, we computed *t*'s for each pair of means with the results shown in Table IV. Scales with nearly identical constant errors—as for example the 4,000 and 9,000 yard scales—are grouped together. The standard error of the difference was computed by taking the figure for the population variance estimated from the interaction (Table III) and *t*-tables were entered with 1 and 100 degrees of freedom. This, of course, yields a more sensitive test of significance than would be achieved with the standard error computed from the 22 items contributing to each mean.

It is immediately apparent from Table IV that there is only one number scale—the 2.5 scale—which is really deviant. For some unexplained reason, this scale gave a negative constant error which was significantly different from the constant error for every other scale—except possibly the 3,000 scale. Constant errors for all scales other than the 2.5 scale do not differ significantly. It is our opinion, incidentally, that the significant interaction in Table IV can also be traced to the 2.5-scale data. Although this scale gave the largest average *negative* constant error, note in Table II that subject H had his second-highest *positive* constant error with the same scale.

TABLE III

ANALYSIS OF VARIANCE OF THE DATA IN TABLE II

Source of Variance	Sums of Squares	Degrees of Freedom	Estimated Population Variance
Between Subject Means	70.70	10	7.07
Between Scale Means	221.66	10	22.17**
Interaction	631.17	100	6.31*
Between Instruments Within Cells	436.95	121	3.61
Total	1,360.48	241	

* Significant at the 1% level

** Significant at the 0.1% level when tested against the interaction

Variable Errors for the 11 Scales

Variable errors for the same data are shown in Table V and the analysis of variance of these data is shown in Table VI. As in the case of the constant errors there are some marked differences among the various scales and these differences are highly significant. Comparisons between pairs of means for the various scales are shown in Table VII which gives *t*-values and levels of significance for all worthwhile comparisons. As in Table IV, scales with nearly identical means are grouped together.

Table VII shows that differences between the mean errors for the 10,000, 1,000, 2,000, 5,000, 4,000, and 8,000 scales are not significantly different. The 9,000 scale, however, is significantly worse than the 10,000 scale at the one percent level, and worse than the 1,000 scale at the two percent level of confidence. The null hypothesis is acceptable for the differences between the 9,000 scale and the 2,000, 5,000, 4,000, and 8,000 scales.

The 3,000 and 6,000 scales are significantly worse than the 10,000 and 1,000 scales at the one percent level of confidence. They are not significantly different from the 2,000, 5,000, 4,000, 8,000, and 9,000 scales.

The 7,000 scale is worse than the 10,000 and 1,000 scales at the one percent level; it is worse than the 2,000, 5,000, 4,000, and 8,000 scales at between the 5 and 10 percent levels; and is not significantly worse than the 9,000, 3,000, and 6,000 scales.

The 2.5 scale is significantly worse than every other scale. In some

TABLE IV

THE *t* RATIOS FOR THE SIGNIFICANCE OF DIFFERENCES BETWEEN MEAN CONSTANT ERRORS SHOWN IN TABLE II

Entries in parentheses are the levels of significance attached to each *t*. Scales with nearly identical constant errors are grouped together. For every comparison the scale at the top of the table has a more positive constant error than the one with which it is compared along the side of the table.

	4,000 9,000	8,000	6,000	10,000	5,000 1,000	2,000	7,000	3,000
8,000	——	——						
6,000	——	——	——					
10,000	——	——	——	——				
5,000–1,000	——	——	——	——	——			
2,000	1.68(10%)	——	——	——	——	——		
7,000	1.82(10%)	1.66(10%)	——	——	——	——	——	
3,000	2.18(5%)	2.02(5%)	1.84(10%)	——	——	——	——	——
2.5	4.52(0.1%)	4.36(0.1%)	4.17(0.1%)	3.61(0.1%)	3.51(0.1%)	2.84(1%)	2.69(1%)	2.34(3%)

$$S.E._{Diff} = \left[\frac{6.31}{22} + \frac{6.31}{22}\right]^{\frac{1}{2}} = 0.757.$$

cases, the significance is better than the 0.1 percent level. This is true, for example, for the comparison of the 2.5-mile scale with the 5,000-yard scale, which is especially interesting because both kinds of scales have been used on modern radars. As was pointed out earlier, the 2.5-mile scale is equivalent to the 5,000-yard scale because a nautical mile is equal to 2,000 yards. Despite this numerical equivalence, our data show that variable errors of interpolation are 1.6 times greater for the 2.5-mile scale as compared with the 5,000-yard scale. This is a striking illustration of the psychological non-equivalence of two numerically-equivalent scales. It is worth noting in passing that the variable errors with the 2.5-mile scale are nearly twice as large as those for the best scale—the 10,000-yard scale.

TABLE V

STANDARD DEVIATIONS OF RELATIVE ERRORS OF ESTIMATION FOR EACH SUBJECT AND EACH CONDITION

The entry in each cell is based on 25 estimates. The standard deviations are computed around the mean error for each cell as given in Table II.

Subject	Instrument	Number Scale											Mean
		1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000	10,000	2.5	
A	1	3.96	2.74	12.90	3.35	4.05	5.88	3.89	3.29	6.95	3.42	2.74	4.49
	2	3.95	3.36	3.68	4.71	3.48	4.51	3.30	6.58	4.28	3.19	4.48	
B	1	3.36	2.67	7.47	6.18	2.43	2.69	3.66	6.01	6.00	3.34	6.62	4.74
	2	4.72	3.86	7.25	3.01	2.86	5.42	4.37	3.16	4.88	3.66	10.64	
C	1	3.07	6.88	5.94	3.15	3.07	4.53	5.18	5.85	5.03	3.50	7.08	4.50
	2	3.92	2.32	2.95	3.50	5.29	3.49	7.09	3.23	4.53	4.30	5.09	
D	1	2.95	5.71	2.75	3.55	2.96	7.94	3.93	6.61	2.96	3.27	7.20	4.82
	2	2.99	5.41	7.12	6.90	7.84	4.77	5.66	3.48	3.98	2.53	5.57	
E	1	12.91	6.02	6.03	4.11	6.54	6.42	12.84	5.74	5.54	4.61	8.28	7.07
	2	3.36	5.16	6.77	4.85	6.57	6.47	12.53	7.77	5.77	4.50	12.76	
F	1	4.76	5.42	14.59	6.37	6.45	6.46	5.63	5.02	7.00	5.12	17.39	7.25
	2	5.54	5.19	8.01	5.91	5.67	5.84	7.39	5.39	12.03	5.29	9.05	
G	1	4.25	2.85	3.82	3.08	3.93	5.38	2.81	3.60	4.08	4.16	5.55	4.76
	2	3.15	3.04	4.76	3.98	10.42	6.22	5.08	7.24	6.23	3.01	8.14	
H	1	2.62	6.45	7.13	7.80	5.34	8.94	3.68	3.40	9.98	9.89	6.73	5.88
	2	2.76	4.44	3.91	4.64	4.30	9.16	5.86	5.87	3.96	3.97	8.60	
I	1	2.73	3.04	2.37	2.71	4.12	4.61	3.86	3.07	5.32	2.67	11.84	3.98
	2	3.50	10.06	2.60	2.79	2.57	3.22	5.24	2.44	2.70	2.30	3.78	
J	1	2.73	3.66	3.11	6.16	4.49	5.56	8.12	3.70	5.12	4.26	5.53	4.80
	2	3.95	4.40	4.95	4.75	4.57	5.53	4.74	5.74	3.57	2.93	8.01	
K	1	2.73	5.58	3.04	4.64	2.64	7.33	4.20	2.87	6.75	3.15	2.89	4.28
	2	3.85	3.57	3.03	9.22	2.27	2.96	8.89	4.85	3.91	2.59	3.10	
Mean		3.99	4.63	5.64	4.79	4.63	5.61	5.82	4.77	5.48	3.89	7.32	5.14

Mean (Instrument 1) = 5.21; Mean (Instrument 2) = 5.08.

TABLE VI

ANALYSIS OF VARIANCE OF THE DATA IN TABLE V

Source of Variance	Sums of Squares	Degrees of Freedom	Estimated Population Variance
Between Subject Means	268.15	10	26.81*
Between Scale Means	208.19	10	20.82*
Subject-Scale Interaction	463.05	100	4.63
Between Instruments Within Cells	491.29	121	4.06
Total	1,430.68	241	

* Significant at the 0.1% level

Analyses with Other Methods of Measuring Errors

To check on the stability of these findings, we recalculated the analyses presented in Tables V and VI using different methods of computing errors. For example, we calculated standard deviations of relative errors of estimation around the mean constant errors shown along the bottom of Table II instead of around the mean constant errors for each cell of Table II. (The latter method, of course, is the one we used to get the values in Table V.) Then we also calculated average errors of estimation (average deviations) for the data when they are uncorrected for the constant errors of calibration shown in Fig. 3. The results of these alternative analyses are gratifyingly consistent. There are some slight shifts in the rank-orders of the various scales, but the shifts are never more than two rank-order positions. No matter how errors are defined, the 1,000, 10,000, 2,000, and 5,000 scales are always the best four; the 3,000, 6,000, 7,000, and 2.5 scales are always the worst four; and the 4,000, 8,000, and 9,000 scales always rank in the middle.

DISCUSSION

The primary objective of this research was to obtain information on the accuracy of visual interpolation when various number systems are assigned to scale markers. Significant differences have been demonstrated between numerical scales in this regard, but for purposes of practical scale design the information obtained

TABLE VII

THE *t* RATIOS FOR THE SIGNIFICANCE OF DIFFERENCES BETWEEN MEAN STANDARD DEVIATIONS SHOWN IN TABLE V

Entries in parentheses are the levels of significance attached to each *t*. Scales with nearly identical mean standard deviations are grouped together. For every comparison the scale at the top of the table is better than the one with which it is compared along the side of the table.

	10,000	1,000	2,000 5,000	4,000 8,000	9,000	3,000 6,000	7,000
1,000	———						
2,000–5,000	———	———					
4,000–8,000	———	———	———				
9,000	2.62(1%)	2.45(2%)	———	———			
3,000–6,000	2.83(1%)	2.66(1%)	———	———	———		
7,000	3.17(1%)	3.01(1%)	1.96(10%)	1.69(10%)	———	———	
2.5	5.64(0.1%)	5.48(0.1%)	4.42(0.1%)	4.16(0.1%)	3.03(1%)	2.76(1%)	2.47(2%)

$$S.E._{Diff} = \left[\frac{4.06}{22} + \frac{4.06}{22}\right]^{\frac{1}{2}} = 0.608.$$

from this experiment must be combined with the data of other, related studies. We have shown that numerical scale units of 1,000 and 10,000 yards give maximum accuracy of visual interpolation. From this we cannot generalize that all scales should have numerical units of 1,000 or 10,000 since both the physical size and the total numerical range of the scale also have to be considered. Leyzorek (4), for example, has shown that best relative accuracy of interpolation occurs in scale intervals ½ in. in size. Thus for some sizes of display it might be better to use a 2,000 or 5,000 scale.

Still to be explored is the possibility that number system and physical scale size may interact in such a way as to further limit the generality of these findings. The fact that this experiment studied only a particular series of numbers with a specific unit of length (yard) also restricts the degree of generalization possible. These limitations grew out of the nature of the equipment used.

A number of interesting questions are raised by the line of investigation pursued in this experiment. Superficially, at least, the visual interpolation problem remained the same for the subject regardless of the numerical scale with which he worked. But the difference in performance, for example, between the numerically-equivalent 5,000-yard and 2.5-mile scales shows that the psychological problem was not the same, at least for some subjects, with these two scales. The question arises: In what way did the numerical scale influence the judgments made by the subjects?

The process of visual interpolation probably involves setting up a series of 'anchor points' in the interval between scale markers. For example, the subject may visually divide the scale interval into halves, quarters, or thirds, and may then subdivide these fractions further. It is largely in terms of such fractional units or anchor points that the subject makes his estimate of the position of a point, later converting the fractional part of the interval corresponding to the point's position into its numerical equivalent in the scale being used. We then ask: (a) Does the subject use one fractionation technique all the time and convert his fractions into the scale in use at the time, or does he use a different set of anchor points, and fractions for each numerical scale? (b) Is there a true semantic or mathematical difference between the numerical scales which renders fractionation or conversion of fractions more difficult in one number system than in another? (c) Is there some optimal fractional unit or combination of units which will give maximum accuracy in visual interpolation? For the answers to these questions, new experiments must presumably be devised.

Why any one numerical scale was superior to the others is a question which remains unanswered by this experiment, but some logical speculation on this subject is defensible. The superiority of the 1,000- and 10,000-yard scales seems clearly based on our almost universal acceptance and use of a decimal system of numbers. Vernon (5) has shown that whole numbers in the decimal system, such as 1,000, 100, 10, gave the fewest errors while scale intervals in steps of four gave the most errors. Scale divisions in steps of two were almost as good as those in the decimal units. In general, changing the scale number by a factor of 10 did not significantly alter its efficiency. Another finding of Vernon's study was that decimal fractions such as .01, .1, .2, were much harder to use than were whole numbers like 1, 10, 2, 20. This is consistent with our finding that the 2.5-mile scale gave the poorest results.

Yet to be determined is the basis for the superiority of the decimal scale unit. Certainly, familiarity is one important factor, but the mental process involved in the fractionation of numerical scale units needs to be studied for the final answer.

Summary

This experiment was designed to get information on the relationship between accuracy of visual interpolation and the numerical system assigned to scale markers. The visual problem presented to the subject contained a number of short arcs (called localization points) scattered in and concentric with a pattern of four equally-spaced, concentric circles which served as scale markers. The subject's task was to estimate the position of each localization point in the field relative to the center of the display and to the scale markers between which it fell.

Eleven different numerical scale units were tested: 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000 yards, and 2.5 miles. For each set of estimates comprising one trial the subjects used the number assigned for that trial, i.e., in one trial the subject made his estimates in terms of a 2,000-yard scale interval unit while in another trial he used a 4,000-yard unit, and so on. The experimental trials were arranged in an 11 × 11 factorial design with one replication.

Relative errors of estimation were analyzed into two components, constant errors and variable errors, the former being expressed by the mean, the latter by standard deviations around the mean. Differences between the constant errors and variable errors for the different scales were found to be highly significant. In the case of the constant errors, the significance was due entirely to one deviant scale—the 2.5 scale—which had a large negative constant error as compared with the other scales. Constant errors for the other 10 scales did not differ significantly.

There were several highly significant differences among the scales in terms of error variability. Ranked in order from best to worst (smallest to largest variable errors) the scales are: 10,000, 1,000, 2,000–5,000 (equal), 8,000, 4,000, 9,000, 6,000, 3,000, 7,000, and 2.5. The worst scale (2.5 miles) gave errors about twice as large as those obtained with the best scale (10,000 yards).

(Manuscript received for immediate publication July 10, 1950)

References

1. Bartlett, N. R., Reed, J. D., & Duvoisin, G. Estimations of distance on polar coordinate plots as a function of the scale used. *J. gen. Psychol.*, 1949, 41, 47–65.
2. Chapanis, A. Accuracy of interpolation between scale markers as a function of scale interval number. *Amer. Psychologist*, 1947, 2, 346.
3. Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied experimental psychology*. New York: Wiley, 1949.
4. Leyzorek, M. Accuracy of visual interpolation between circular scale markers as a function of the separation between markers. *J. exp. Psychol.*, 1949, 39, 270–279.
5. Vernon, M. B. Scale and dial reading. Med. Res. Counc. Unit in Appl. Psychol., Univ. Cambridge, England, Rep. No. APU49, June, 1946.
6. Williams, S. B., & Bartlett, N. R. Visibility on cathode-ray tube screens: problems and methods. *J. Psychol.*, 1948, 25, 401–417.
7. Williams, S. B., Bartlett, N. R., & King, E. Visibility on cathode-ray tube screens: screen brightness. *J. Psychol.*, 1948, 25, 455–466.
8. Yule, G. W. On reading a scale. *J. roy. statist. Soc.*, 1927, 90, 570–587.

THE VISUAL DISCRIMINATION OF GEOMETRIC FORMS *

BY ROLAND CARL CASPERSON

The Johns Hopkins University **

INTRODUCTION

Form discrimination as it is used here means the identification of a particular geometric form as being different from all other possible ones. Practically, the total number of forms in any experiment is usually small so that discrimination involves a choice of one among the several forms presented. The discriminability of a form is usually measured by frequency or percent of correct reports as a function of some variable. Those variables most often investigated in the literature have been size, intensity of illumination and position in the retinal field. Regardless of the variable chosen, the relative discriminability of a particular form is essentially unaffected.

Theoretical Background

Empiricists predict that a familiar form will be easier to perceive or discriminate. Although it has been demonstrated (1, 5, 6, 7, 9, 12) that experience is related to the discrimination of form, it is practically impossible to apply the Empirical theory to geometric forms. Most geometric forms are already so familiar to the average *O* that no amount of experimental practice would produce the necessary differential in experience.

About the only other systematic treatment of form perception is that of the Gestalt school. Although the Gestalt theory of 'simplicity' or 'good figure' can be applied to geometric figures, most investigations have disproved the configurationist hypotheses. For example, Helson and Fehrer (8) demonstrated the inadequacy of the Gestalt concept of 'simplicity.' According to this concept the circle is the simplest figure and for this reason should be identified more easily than other forms. Actually, Helson and Fehrer found that the circle ranked after the rectangle and triangle in perceptibility. Similar results have been found in other studies (4, 11, 13, 14). Although the order of discriminability for different forms varied in these studies, they all agree that the circle is neither good nor exceptionally poor in discriminability.

A second point which Helson and Fehrer discuss is the Gestalt idea that form is primary in experience. According to the Gestalt theory all experience is formed and therefore form is primary in perception. Several studies in the literature have found this to be untrue (2, 3, 5, 8, 16). Helson and Fehrer found three levels of perception in their experiment which they also considered to be negative evidence for the primacy of form. Their *O*s first reported vague formless light, then a formed object which could not be correctly identified, and finally the correct form. In an attempt to defend the Gestalt hypothesis Wilcox (15) interpreted these findings as showing that the initial perception was formed. Although the light could not be correctly identified as a shape, it appeared as a rounded blur, which fit Wilcox's criterion for form. Such a definition of form leaves little in our experience that is unformed. It also makes any discussion of form perception or discrimination meaningless. If we use the traditional interpretation of discrimination as the distinction and recognition of

* This research has been supported in part under the terms of a contract between Special Devices Center, Office of Naval Research, and The Johns Hopkins University, Contract N5-ori-166, Task Order I. This is report No. 166-I-108, Project Designation No. NR 784-001, under that contract.

** The author is indebted to Dr. W. R. Garner, who aided in the formulation of the problem and saw the research to its completion.

a particular form as different from others, we can dismiss Wilcox's definition of form perception. If a stimulus need only appear as a vague round blur of light to qualify as a form, there is no way of testing the primacy of form in experience.

It is also possible to argue that when form is indistinguishable stimuli are reduced to a roughly circular blur. This, then, is not form but a complete lack of it. Viewed in this way, the circle, instead of being the ideal form as configuration theory maintains, approaches formlessness because it most closely resembles that condition wherein form cannot be discriminated.

Such a conclusion has interesting implications. If we expect stimuli below the recognition threshold to appear round we might expect the frequency of reports for circles, ellipses and other similar figures to be high in any task in which form thresholds are measured. Under these circumstances we should also expect on the basis of chance alone that more circles would be called correctly, and thus falsely have lower thresholds than other forms.

A recent study by Hochberg, Gleitman and MacBride (10) offers one of the few bits of evidence supporting the Gestalt theory of form perception. Using three figures, a circle, rectangle and block cross of equal area, these investigators obtained intensity thresholds for form discrimination. Testing the Gestalt concept of 'simplicity,' which they defined as the ratio of perimeter to area, their results confirmed the hypothesis. The circle had the lowest threshold, while the square and the cross had progressively higher ones.

There are two possible reasons for such excellent results. The first is that these investigators used only three figures. Such a small number of forms can hardly be considered an adequate test of so broad an hypothesis. Second, these authors present no information about the frequency of report for their figures regardless of accuracy. As was mentioned above, a high incidence of circle reports, whether due to preference or the above mentioned conditions, would make the circle appear to be superior. Although we cannot reject configuration as a possible determinant in discrimination, some quantifiable aspect of form should lead us to a little more understanding of the problem. Hochberg, Gleitman and MacBride have made a step in the right direction, but their results fall short of a conclusive determination of the important factors.

Purpose

The specific purpose of this investigation was to determine the discrimination thresholds of six different geometric forms and to relate their relative discriminability to three quantifiable aspects of their construction. To produce variations within the basic forms,[1] five different figures for each form were constructed. These figures, when equal in area, differed in maximum dimension and perimeter. These variations were necessary to demonstrate the effect of perimeter and maximum dimension without contamination by the property of the basic configuration of the six forms. Finally, each set of figures was reproduced at seven different areas in order to measure discrimination thresholds.

Apparatus and Observers

There were 30 different stimulus figures used, consisting of the six basic forms: ellipse, rectangle, triangle, diamond, cross and star. Each of these forms was constructed in five different ways as shown in Fig. 1. Although all of the figures were made to have the same seven areas it was impossible to obtain this equality for the linear dimensions because of the nature of the geometric relations involved. They were constructed, however, so that the maximum dimensions and perimeters for the forms agreed as closely as possible.

Six complete sets of stimulus figures were made for each of seven areas. The legend of

[1] In this paper, the word 'form' will be used to refer to a general class of regular geometrical shapes, for example, star, rectangle, circle or triangle; 'figure' will refer to variations in construction of any form.

Fig. 1 shows the seven areas used and the maximum dimension and perimeter resulting for the smallest set of figures. The linear dimensions for the other sets of figures vary as the square root of the area and can be obtained for each area by using the multiplying factor indicated. The figures were solid black photographic prints on semi-gloss paper; maximum contrast was obtained. They were reduced from a master set of figures which were constructed eight times the size of the largest stimulus area used. In this way any minor error in construction of the original figures was eliminated in reduction. The stimuli were mounted on heavy cards measuring six and one-half in. square. Each card contained nine figures in random order arranged in three rows and columns. The figures were placed two in. apart from center to center, regardless of area. For each area a set of 20 cards was necessary to contain all of the figures.

The stimulus cards were exposed from the far end of an illuminated exposure box measuring 36 in. square. The box was painted flat white on the interior and was lighted by four G. E. Lumiline light bulbs around the four sides at the forward end of the box. The lights were shielded so that no direct light fell on the stimulus cards or on *O*. The illuminance measured on the surface of the stimulus cards was 11.2 foot candles.

O sat 20 feet from the exposure aperture and looked through the open front end of the illuminated box. He was seated comfortably in an adjustable chair and was provided with an arm and chin rest. Although *O* used binocular vision and was allowed to move his head at will, the

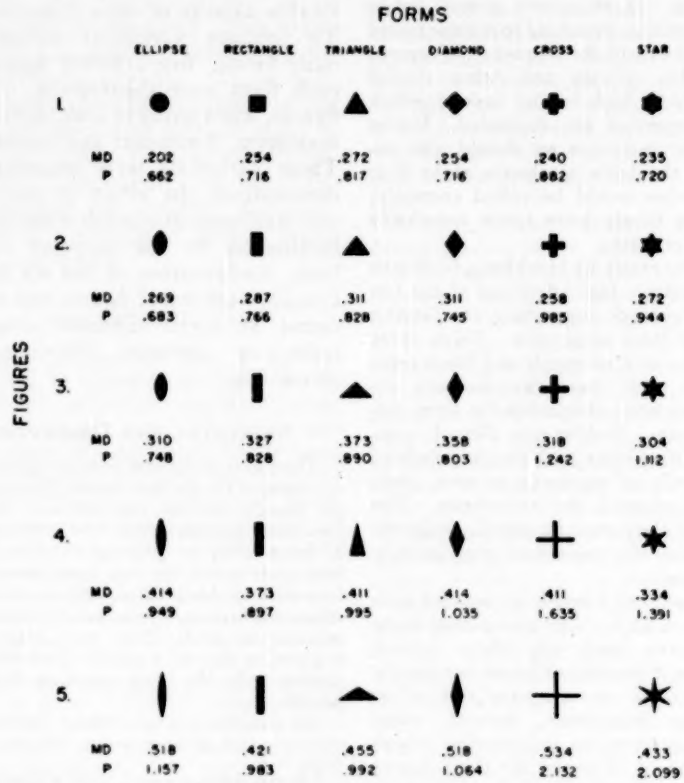


Fig. 1. The experimental forms and figures. Maximum dimension (MD) and perimeter (P) in cms. for the smallest area used are indicated below each figure. These linear measures for the six larger areas can be obtained by multiplying the given values by 1.5, 2, 2.5, 3, 4 and 5. The seven areas used were .032, .072, .128, .200, .288, .512, and .800 cm.^{2}

arm and chin rest positioned his head so that he was directly in front of the exposure aperture at the correct height.

The *Os* were 20 male students ranging from 18 to 27 years of age. All *Os* had normal acuity as measured by the AO test chart; five of them wore corrective lenses. Especial care was taken to avoid astigmatics.

Procedure

Each *O* had seven experimental sessions in each of which he judged a complete set of 30 figures for a particular area 24 times. Each time a figure was repeated it appeared in a different relation to the surrounding figures and in a different orientation. The sessions lasted from one to two hours depending on the size of the figures and the speed with which a subject could make his reports. The order of areas and cards within an area was counterbalanced throughout so that no two *Os* were presented the same order of figures and areas.

A set of written instructions was given to each *O* at his first experimental session and the procedure was discussed with him. He was shown a card containing all of the figures with their proper identifications. He was told only to report in terms of the six basic forms named. It was necessary to add this precaution because of the presence of a circle in the group of ellipses and a square in both the rectangle and diamond group. These distinctions were made quite readily by the *Os* and very few errors of confusion occurred. The *Os* were advised not to spend too much time on any one figure and the method of forced guesses was used.

E sat behind the lighted box where he exposed the cards by means of a chute and lever which positioned the cards until released. *O* started with the upper left figure on a card and read them off from left to right as rapidly as he could without sacrificing accuracy, progressing down the card until he judged all nine figures. No knowledge of results was permitted and no learning was apparent other than familiarization with the procedure.

Results

Because the two major problems of interest in this investigation involve somewhat different treatments of the data, the results are divided into two main sections, with an added discussion on individual differences. Although the results concerning the relative discriminability of different forms might logically seem to take precedence, the data pertaining to discrimination as a function of area, maximum dimension and perimeter are presented first. These data contribute to both questions and therefore reproduction of data is reduced by following this order.

Frequency of reports.—Percent of correct reports, the measure of discriminability, has in all cases been corrected for response probability. Instead of correcting for the theoretical chance prediction based on the number of forms used, the corrections were based solely on the empirical frequency of reports for each of the forms.[2] The correction formula is essentially the same as the one usually used in chance corrections, but the proportion used is determined by the percent of total error responses that are called a particular form.

Table I shows the frequency of responses and the percent of the total possible responses which were made for each form, regardless of accuracy. It can be seen from Table I that the ellipses exceeded all other forms in frequency of reports. The rectangle and triangle are also somewhat higher in frequency than would be predicted by chance, while the diamond, cross and star fall below chance expectancy. This confirms the suspicion raised by the earlier discussion in this paper, and emphasizes the need for a correction based on the frequency of reports.

Discrimination as a Function of Area, Maximum Dimension and Perimeter

The geometric relationships between the variables considered in this

[2] Correction for frequency of report was obtained by $\left(Rc_x = Tx - \frac{W_x}{1 - P_x}\right)$ where: Rc_x = corrected number of right responses for form X; T_x = total number of form X presented; W_x = number of form X wrong; P_x = proportion of total errors made on all forms which were called form X.

TABLE I

FREQUENCY AND PERCENT OF REPORTS FOR EACH FORM REGARDLESS OF ACCURACY
Theoretical chance percent based on the number of forms = 16.7

	Ellipse	Rectangle	Triangle	Diamond	Cross	Star	Total
Frequency	21,607	17,258	19,619	14,591	14,919	12,806	100,800
Percent	21.4	17.1	19.5	14.5	14.8	12.7	100

experiment preclude any possible systematic evaluation of the effects of one variable while the others are held constant. In setting up the experiment the seven different areas were selected to produce functions for all of the forms. The figures within a form on the other hand were constructed to produce variations in maximum dimension and perimeter for these basic forms. It is obvious that these characteristics must vary together and in relation to area. For this reason it is necessary to approach the problem of their influence on discrimination indirectly.

Figs. 2, 3, and 4 show percent correct reports as a function of area, maximum dimension and perimeter, respectively. Although these variables are necessarily related in the construction of the various figures, it can be seen from these plots that the functions do not correlate perfectly. The shape of the curves for the 30 forms and figures differs considerably depending on the variable considered. If any one of the measures investigated in this experiment was the sole determinant of a form's discriminability, all of the functions for that variable should fall on the same curve. The degree to which these functions differ for a particular variable indicates the amount to which some characteristic of form other than that variable contributes in discriminability of the various forms.

In this experiment we are interested in finding that variable which produces discrimination functions that are most alike for all of the figures. This variable then would reduce the variations among functions to a minimum, leaving the least amount of variance between curves to be accounted for by some other property of form.

Inspection of Figs. 2, 3, and 4 shows that the functions differ considerably in the degree to which they approximate a single curve. It can also be seen from these data that the functions differ for the linear measures, i.e., maximum dimension and perimeter, in their points of origin and in their extent. This makes a quantitative comparison of the variance remaining between the functions for the three variables impossible. We must, therefore, obtain some arbitrary value for each variable at which we can compare the functions for all of the forms and figures. The value having the most reliability and having a direct relationship to all of the functions is the 50 percent threshold value obtained by combining the data from all of the functions into a total mean curve[3] (heavy solid line in upper left diagram of Figs. 2, 3, and 4). The threshold value for each variable on this total mean curve (shown as vertical dotted line in each set of functions) is the point about which the

[3] Mean curves were computed for a range including the 25th and 75th percentiles in all cases, based on five arbitrarily chosen values on the abscissa for the linear variables. This was done to account for the differences in the empirical values of each variable used in plotting the functions.

variation between forms and figures is computed. The variance of the individual functions about the total threshold then is a measure of the inadequacy of a variable in predicting the discriminability of these figures. Therefore that variable which produces the least variation among functions contributes most to the discrimination of form.

The total variance for the 30 functions can be partialed into two components, variance between forms and variance within forms. Between-form variance is computed by obtaining the deviations in percent of correct

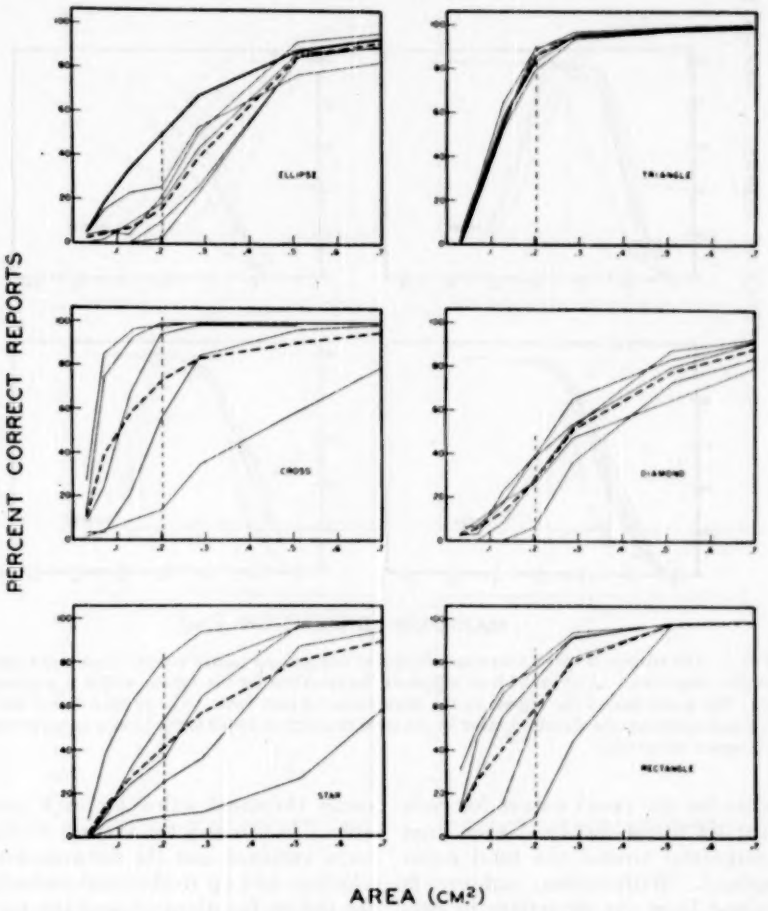


FIG. 2. The relation between area and percent correct reports (each point represents 480 judgments). Light solid lines represent the functions for the figures within a particular form. The heavy dashed line represents the mean curve for each form. The vertical dotted line in each graph represents the threshold value for the total mean curve for all forms (drawn as heavy solid line in upper left graph).

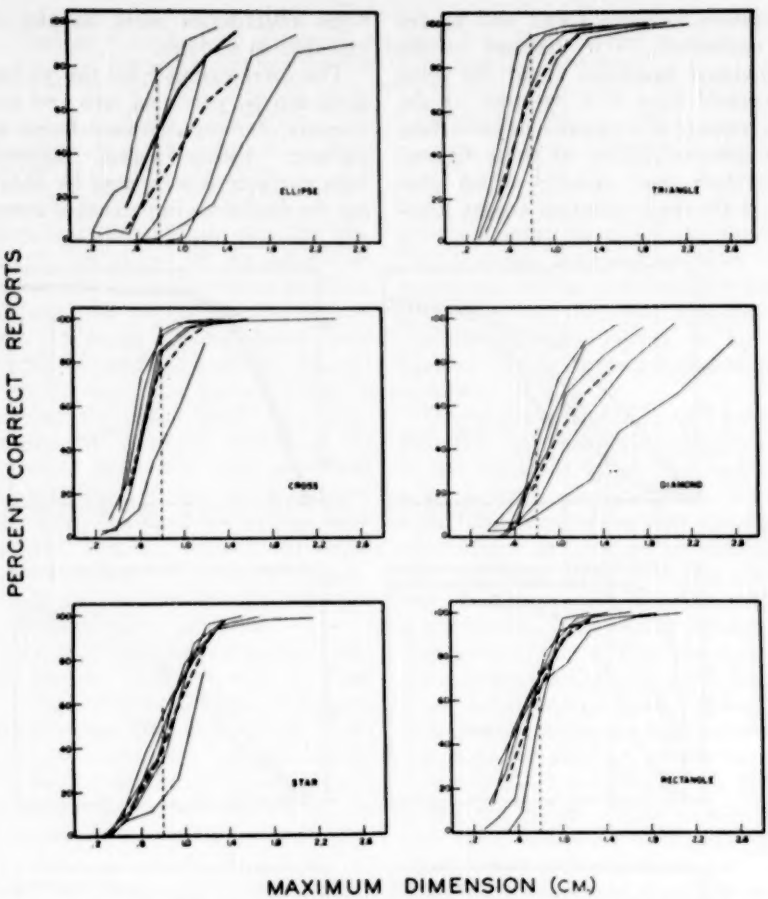


FIG. 3. The relation between maximum dimension and percent correct reports (each point represents 480 judgments). Light solid lines represent the functions for the figures within a particular form. The heavy dashed line represents the mean curve for each form. The vertical dotted line in each graph represents the threshold value for the total mean curve for all forms (drawn as heavy solid line in upper left graph).

reports for the mean curves for each of the six forms (heavy dashed lines in diagrams) around the total mean threshold. Within-form variance is obtained from the deviations in percent of correct reports of the five individual figure curves (light solid lines in illustrations) around their respective mean form curves at the total mean threshold value for each variable. In this way the average within-form variance and the between-form variance add up to the total variation for the 30 functions around the total mean threshold.

Table II shows the amount of variance remaining for the functions of each of the three variables considered,

as measured by the foregoing method. It can be seen from Table II that no one of the variables is successful in minimizing the within-form variance for all six of the basic forms. It appears that area is most successful for the ellipses and triangles. The rectangles and diamonds differ least when maximum dimension is held equal. And both the crosses and stars differ minimally when perimeter is the controlling factor. When we consider the between-form variance it is apparent that maximum dimension does the best job in equating forms for discrimination. Finally the total vari-

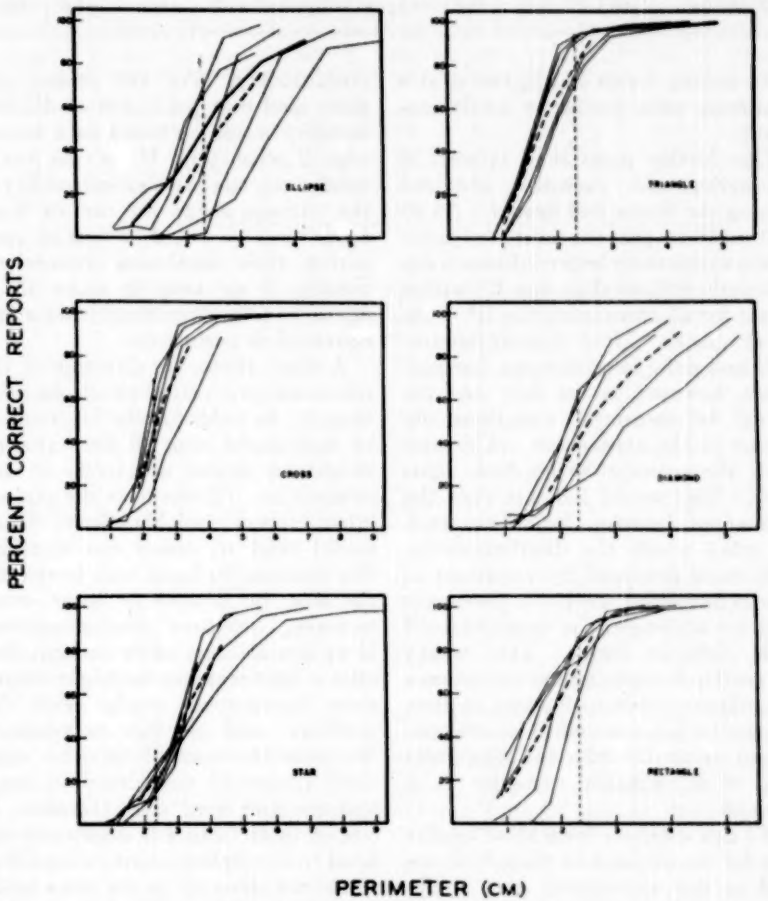


FIG. 4. The relation between perimeter and percent correct reports (each point represents 480 judgments). Light solid lines represent the functions for the figures within a particular form. The heavy dashed line represents the mean curve for each form. The vertical dotted line in each graph represents the threshold value for the total mean curve for all forms (drawn as heavy solid line in upper left graph).

TABLE II

VARIANCE IN PERCENT CORRECT REPORTS REMAINING WHEN FORMS AND FIGURES ARE COMPARED FOR THRESHOLD AMOUNTS OF EACH OF THE THREE VARIABLES: AREA, MAXIMUM DIMENSION AND PERIMETER

Variable	Within-Form Variance							Between-Form Variance	Total Variance
	Ellipse	Rectangle	Triangle	Diamond	Cross	Star	Mean		
Area	68.5	668.7	13.6	136.6	1107.5	590.4	430.9	594	1024.9
Max. Dim.	873.5	46.3	582.8	126.9	373.4	90.4	348.9	399.6	748.5
Perimeter	766.3	107.9	95.5	243.3	246.9	61.2	253.5	436.8	690.3

ance among forms and figures is at a minimum when perimeter is held constant.

One further point is of interest in comparing the variances obtained among the forms and figures. An *F*-test indicates that the estimated population variance for between forms is significantly greater than that for within forms for all three variables ($P < .01$ in all cases). More directly observable and more important for our purposes, however, is the fact that the actual between-form variations obtained in the experiment are greater than the average within-form variance. This would indicate that the differences between forms are real. In other words the discriminability differences produced by variations in the construction within a particular form are not as great as those obtained with different forms. This means we must acknowledge the importance of configuration in such a task as form discrimination, even when we are concerned primarily with the manipulation of quantifiable variables as a method.

We can conclude from these results that for the ellipse and triangle forms used in this experiment area is the important factor in discrimination. Knowing the maximum dimension of rectangular and diamond shaped forms would appear to contribute most to our prediction of their discriminability. For the crosses and stars used best prediction of discriminability would be based on a knowledge of perimeter. If our task was to predict the relative discriminability of the average shape for the six basic forms used we would do best by comparing their maximum dimensions. Finally, if we were to make all 30 figures equally discriminable we would equate their perimeters.

A word about the direction of the relationships involved should be mentioned. In order for the functions to be meaningful none of the variables should be related negatively to discrimination. If this were the case we might expect variable effects which would tend to cancel one another. For instance we know that increasing the size, be it area or some other measure, increases discriminability. If we now assume, safely enough, that area is that measure we might obtain some inconsistent results with the methods used in this experiment. We know that each of the other variables concerned must vary in some manner with area and, therefore, if one of these factors is negatively related to discrimination its relationship would not show up in the plots used. In fact the effects of such a negative relationship might serve to reduce or remove the effects of any positive relationship present.

Table III shows the percent of cor-

rect reports for each form and figure as a function of area. In Fig. 1 we saw that maximum dimension increases with the number assigned to the figure within a particular form. This is also the case for perimeter except in the case of two triangles (4 and 5) in which there is a reversal of the relationships. At any rate an inspection of Table III reveals that all three variables are related positively to discrimination for the rectangles,

TABLE III

PERCENT OF CORRECT REPORTS FOR ALL OF THE GEOMETRIC FIGURES AS A FUNCTION OF AREA, INCLUDING MEANS FOR THE SIX BASIC FORMS AND FOR THE SEVEN AREAS

Area:	.032	.072	.128	.200	2.88	.512	.800	Mean
Ellipse								
1	4.2	15.0	22.7	25.4	53.2	77.1	85.6	40.5
2	3.7	5.1	3.9	20.2	51.1	92.1	98.2	39.2
3	0.0	0.0	0.0	9.9	26.7	84.6	96.0	31.0
4	0.0	0.0	0.0	2.0	23.5	85.0	95.2	29.4
5	0.0	1.4	8.8	17.8	43.7	87.1	91.2	35.7
Mean	1.6	4.3	7.1	15.1	39.6	85.2	93.2	35.2
Rectangle								
1	0.0	0.0	0.0	9.2	44.7	97.4	100.0	35.9
2	1.2	7.0	15.7	54.4	82.3	97.6	99.8	51.1
3	11.5	27.6	55.2	71.2	91.2	98.3	100.0	65.0
4	12.7	46.1	65.5	79.8	93.9	97.9	99.6	70.8
5	30.7	51.1	69.0	76.8	91.8	98.3	99.8	73.9
Mean	11.2	26.4	41.1	58.3	80.8	97.9	99.8	59.3
Triangle								
1	0.0	23.9	55.1	89.3	94.9	98.5	99.8	65.9
2	2.0	22.5	56.9	86.0	95.1	97.8	98.9	65.6
3	4.4	27.7	65.0	88.8	96.7	98.7	100.0	68.7
4	0.0	24.4	57.1	83.4	95.8	98.0	99.8	65.5
5	0.0	20.0	53.6	79.3	93.5	97.6	99.3	63.4
Mean	1.3	23.7	57.5	85.4	95.2	98.1	99.6	65.8
Diamond								
1	0.0	0.0	0.0	6.3	36.3	72.7	90.3	29.4
2	0.0	0.0	8.2	28.1	52.7	87.4	97.5	39.1
3	2.6	3.0	22.9	38.5	65.5	83.4	96.3	44.6
4	6.5	6.6	17.3	38.0	53.9	79.8	97.5	42.8
5	2.6	9.0	18.3	26.4	47.7	64.8	89.6	36.9
Mean	2.3	3.7	13.4	27.4	51.2	77.6	94.2	38.6
Cross								
1	1.4	4.4	8.4	13.9	34.4	59.4	89.4	39.2
2	3.3	3.8	20.6	55.8	84.4	96.0	99.5	51.9
3	8.1	24.6	66.2	94.0	98.2	98.5	100.0	70.0
4	12.4	74.1	93.2	98.9	99.6	99.1	100.0	82.5
5	26.6	85.1	96.2	98.5	99.3	100.0	100.0	86.5
Mean	10.4	38.4	56.9	72.2	83.2	90.6	97.8	64.2
Star								
1	0.0	2.2	7.1	9.5	11.2	26.2	74.6	18.7
2	0.0	7.8	13.0	24.8	36.0	87.3	97.3	38.0
3	0.0	12.8	24.6	35.6	64.2	96.1	99.5	47.5
4	0.7	10.2	32.1	52.3	78.2	95.7	99.0	52.6
5	8.7	38.4	61.5	80.0	93.8	97.7	98.3	68.3
Mean	1.9	14.3	27.7	40.4	56.7	80.6	93.8	45.0
Total mean	4.8	18.5	33.9	49.8	67.8	88.3	96.4	51.4

crosses and stars. For the other three forms the relationship is indeterminate for both maximum dimension and perimeter. In no case is there a negative relationship between any of the variables and percent correct reports.

It would appear then that the knowledge we have about discrimination can logically be extended to the direction of the relationship as well as the relative prediction value of each. Increasing that variable which is most successful in minimizing the variance between figures and forms as shown in Table II will be most successful in improving the discrimination of that particular figure or form.

Relative Discriminability of Forms

The relative discriminability of the six forms used in this experiment can be evaluated in several ways. In keeping with the methods most often used in the literature two comparisons have been made. The first comparison actually provides us with three measures for these data. In a comparison of the discrimination thresholds for the different forms it is necessary to consider each of the three variables measured. Since area, maximum dimension and perimeter were varied we can obtain the threshold for each form in terms of each one of these variables. The second comparison is for the total percent of correct reports for each form regardless of the variable conditions.

Table IV summarizes these comparisons for the mean and best figure curves of each form for the 50 percent thresholds and for total percent correct judgments. An inspection of Table IV reveals several things of interest. First, in none of the comparisons does the ellipse rank as a 'good figure.' In only one case (i.e., perimeter) does it rise above last place in the rank of discriminability and then it attains only a rank of fifth out of the six possible forms. Although the forms shift slightly in their rank, depending on the measure used we can see that the triangle, cross and rectangle consistently maintain their ranks within the first three positions, while the star, diamond, and ellipse occupy the lower three positions, regardless of the measure.

Comparison of best figures.—Because a circle, the best figure of the forms called ellipses, might suffer unduly by being grouped with other ellipses, a comparison was made of the best figures within each form. Table IV also shows a comparison of the best figure for each form for the four conditions. It is evident from these comparisons that essentially the same results are obtained. Although the best ellipse (circle) rises to a rank of three in the case of perimeter measures, it still remains in the lower half of the distribution for three of the four comparisons, and retains its last place in two instances. These results are consistent with the results cited in the literature on size, area and visual field, except for the recent study by Hochberg, Gleitman and MacBride (10). When we consider the forms that they used, namely, the circle, square and block cross, we can only confirm their results. Comparing the ranks of these particular figures as obtained in this experiment the circle also appears to be easiest to identify. The square is next in discriminability, while the block cross fares worst.

Although this experiment confirms the results of these authors, it reverses the conclusions they draw from their results, and in fact makes clear why their results fit so nicely into their theory. The circle, which is the best possible ellipse form, is compared with the square and block cross, which are

TABLE IV

RELATIVE DISCRIMINABILITY OF SIX GEOMETRIC FORMS DETERMINED BY THE AMOUNT OF AREA (A), MAXIMUM DIMENSION (MD) AND PERIMETER (P) NECESSARY TO PRODUCE 50 PERCENT CORRECT REPORTS, AND BY THE TOTAL PERCENT CORRECT REPORTS REGARDLESS OF AREA

Data are given for mean form and for best figure within a form classification. Threshold values for the mean and best figure forms can be obtained from interpolation on the curves presented in Figs. 2, 3, and 4. Total percent correct reports for the respective forms can be obtained from Column eight in Table III.

Forms	Mean Form				Best Figure for Each Form			
	50% Thresholds			Total Percent Correct	50% Thresholds			Total Percent Correct
	A	MD	P		A	MD	P	
Ellipse	.338	1.080	2.62	35.2	.275	0.595	1.96	40.5
Rectangle	.165	0.690	1.84	59.4	.070	0.592	1.43	73.9
Triangle	.117	0.670	1.69	65.8	.105	0.522	1.37	68.7
Diamond	.285	1.040	2.54	38.6	.238	0.860	2.18	44.6
Cross	.105	0.635	2.43	64.2	.046	0.335	2.13	86.5
Star	.255	0.870	3.22	45.0	.099	0.760	3.09	68.3

both the worst figures for their respective forms. Evidence of this fact can be obtained by comparing the mean percent correct reports for each of these figures in Table III (these three figures are number 1 in their respective form categories). It can be seen that the circle is best, receiving 40.5 percent correct reports, while the square and block cross each received 35.9 and 30.2 percent, respectively. Instead of confirming the conclusions of these authors then, these results are more important as evidence for the inadequacy of their test of the 'simplicity' hypothesis. Further, the results of this experiment demonstrate the inadequacy of the Gestalt theory of 'simplicity' or 'good figure' in the discrimination of form. From these data, it would appear that increasing the maximum dimension and perimeter, or the complexity of a form, in most cases increases the probability of its being seen.

Individual differences.—One point that is often neglected in the literature on form perception is the problem of individual differences. This problem is not only crucial to any principles or formulations concerned with form discrimination, but it is very important in the simple interpretation of results. When we say that one form can be identified more easily than another we imply that this is a stable characteristic of a particular form and holds for all individuals. Such an implication may not be justified. For this reason an analysis of the variations in percent correct reports between individuals and between the six basic forms was made. If we maintain that the relative discriminability of a form is a property of that form and not a characteristic of a particular individual at some given time, we would expect the form differences to be significant in such an analysis.

Table V summarizes the results of an analysis of the variance contributed by the 20 individual *Os* in this experiment as compared with the variance due to form differences for the six basic forms. *F*-tests show that the variance due to form and individual differences are both extremely

TABLE V

ANALYSIS OF THE VARIANCE CONTRIBUTED BY FORM AND INDIVIDUAL DIFFERENCES
Number of forms = 6; number of individual *Os* = 20.

Source of Variance	Sums of Squares	Degrees of Freedom	Estimated Pop. Variance
Between forms	1,319,945.08	5	263,989.02
Between *Os*	793,176.81	19	41,746.15
Remainder	178,399.59	95	1,877.89
Total	2,291,521.48	119	

significant ($P < .01$), although the significance is far greater for form differences. An inspection of column 2 in Table V presents these findings and shows that the sum of squared deviations obtained between forms is much greater than the sum of squared deviations betweeen individuals. These data then verify the earlier conclusion that there are real differences in the discriminability of the six forms used in this experiment. Further, it is clear that if there are any inconsistencies between *Os* in these functions, these inconsistencies are very small compared to the over-all effect of form.

SUMMARY AND CONCLUSIONS

Discrimination data were obtained for 30 solid geometric figures consisting of five variations in the construction of each of six basic forms: an ellipse, diamond, triangle, rectangle, cross and star. Percent of correct reports, corrected for response frequency, was plotted as a function of area, maximum dimension and perimeter. The effectiveness of each variable in predicting discriminability for the different forms and figures was determined by comparing the variance remaining between figures having equal threshold amounts of each of the three variables. That variable which produced the least variance between functions was considered the best predictor of discriminability.

Results

1. Area was found to be the best measure of discriminability for ellipses and triangles.

2. Maximum dimension best predicted discriminability for rectangles and diamonds.

3. Perimeter was the best predictor for stars and crosses.

4. As a group the six basic forms can be evaluated best for discriminability when their maximum dimensions are known.

5. Prediction of discriminability for the total of 30 forms and figures was best accomplished on the basis of their perimeters.

6. Comparison of the relative discriminability of the six forms used in this experiment in terms of thresholds and total percent of correct reports confirms earlier findings that the circle and elliptical shapes in general are not very easy to identify.

7. These results indicate that the Gestalt principle of 'simplicity' is inadequate as a predictor of the relative discriminability of ellipses, rectangles, triangles, diamonds, crosses, and stars such as those used in this experiment.

8. A comparison of the variance contributed by the 20 *Os* with the variance due to form differences substantiates the hypothesis that forms do differ in their discriminability and that differences among individuals making the discrimination are small compared to these form differences.

(Manuscript received for immediate publication July 17, 1950)

REFERENCES

1. BRALEY, K. W. The influence of past experience in visual perception. *J. exp. Psychol.*, 1933, **16**, 613–643.

2. Bridgen, R. L. A tachistoscopic study of the differentiation of perception. *Psychol. Monogr.*, 1933, **44**, No. 197.
3. Brown, R. H., & Niven, J. I. The relation between the foveal intensity threshold and length of an illuminated slit. *J. exp. Psychol.*, 1944, **34**, 464–476.
4. Collier, R. M. An experimental study of form perception in indirect vision. *J. comp. Psychol.*, 1931, **11**, 281–290.
5. Dickinson, C. A. Experience and visual perception. *Amer. J. Psychol.*, 1926, **37**, 330–344.
6. Djang, S. The role of past experience in the visual apprehension of masked forms. *J. exp. Psychol.*, 1937, **20**, 29–59.
7. Duncker, K. The influence of past experience upon perceptual properties. *Amer. J. Psychol.*, 1939, **52**, 255–265.
8. Helson, H., & Fehrer, E. V. The role of form in perception. *Amer. J. Psychol.*, 1932, **44**, 79–102.
9. Henle, M. An experimental investigation of past experience as a determinant of visual form perception. *J. exp. Psychol.*, 1942, **30**, 1–21.
10. Hochberg, J. E., Gleitman, H., & MacBride, P. D. Visual threshold as a function of simplicity of form. *Amer. Psychologist*, 1948, **3**, 341–342 (Abstr.).
11. Kleitman, N., & Blier, Z. A. Color and form discrimination in the periphery of the retina. *Amer. J. Physiol.*, 1928, **85**, 178–190.
12. Moore, M. G. Gestalt vs. experience. *Amer. J. Psychol.*, 1930, **42**, 453–455.
13. Munn, N. L., & Geil, G. M. A note on peripheral form discrimination. *J. gen. Psychol.*, 1931, **5**, 78–88.
14. Whitmer, C. A. Peripheral form discrimination under dark adaptation. *J. gen. Psychol.*, 1933, **9**, 405–419.
15. Wilcox, W. Helson and Fehrer on "The role of form in perception." *Amer. J. Psychol.*, 1932, **44**, 578–580.
16. Zigler, M. J., Cook, B., Miller, D., & Wemple, L. The perception of form in peripheral vision. *Amer. J. Psychol.*, 1930, **42**, 246–259.

FIGURE–GROUND REVERSAL AS A FUNCTION OF VISUAL SATIATION [1]

BY JULIAN E. HOCHBERG

Cornell University

According to the theory proposed by Köhler and Wallach to explain the visual effects obtained after prolonged fixation of an 'inspection' pattern (**2**), the perception of a figure entails the flow of a 'figural current' through the cortical medium, and the raised 'resistance' (presumably some phenomenon of electrotonus) of the medium resulting from the current flow is assumed to be responsible for the figural after-effects obtained. Köhler suggests (**1**, p. 72) that some such process of 'satiation' is also responsible for the reversal of ambiguous figures:

If it is assumed that the cortical region isomorphically corresponding to the perceived figure is the seat of the highest concentration of cortical current flow (**2**, p. 328), then whichever portion of an ambiguous pattern is first seen as figure should satiate its corresponding cortical region at a faster rate than do the other portions of the field. When the satiation-induced resistance opposing current flow in this region becomes sufficiently high a restructuring of the percept should result, with the alternate portion of the pattern then being perceived as figure until its satiation results in further reversal, etc., until, perhaps, the entire region becomes extremely satiated.[2]

If satiation is at least one of the factors responsible for the reversal of ambiguous figures, it should be possible to interfere with the process by imposing a differential satiation upon portions of the cortical region in which the figural currents of the ambiguous figure are to flow. Thus, if a figure, such as Fig. 1c, corresponding to one of the crosses of the reversible figure, Fig. 1a, is presented for prolonged inspection, the cortical region occupied by its isomorphic figure current would attain a raised resistance; on subsequent fixation of the reversible Fig. 1a, the cross falling on a less satiated cortical region should be more easily perceived as figure. Köhler and Wallach hold (**2**, p. 331) that the figural currents isomorphic to the perception of a black figure on white background are identical with those of a white figure on black ground *except that the direction of flow is opposite in the two cases;* and that the satiation-induced resistance (**2**, p. 332) is *non-directional* (i.e., a region satiated by a current of one direction will equally oppose current flow in the same or in the opposite direction). Consequently, it should make no difference whether the satiating figure and the coincident portion of the reversible test figure are of the same 'color' (i.e., both black on a white ground, or both white on a black ground) or are of

[1] This paper reports a modified repetition of a portion of a dissertation submitted to the Department of Psychology of the University of California at Berkeley in partial fulfillment of the requirements for the Ph.D. degree, sponsored by Professors Edward C. Tolman, Egon Brunswik, and Franklin M. Henry, 1949.

[2] A rigorous description of the process of figure-ground reversal in terms of the model proposed by Köhler and Wallach would probably entail consideration of the redistribution of currents in the ground as well as in the figure, and a knowledge of characteristics of current and medium as yet unknown; the hypothesis as presented by Köhler, however, seems sufficiently clear for the purposes of the problem here considered.

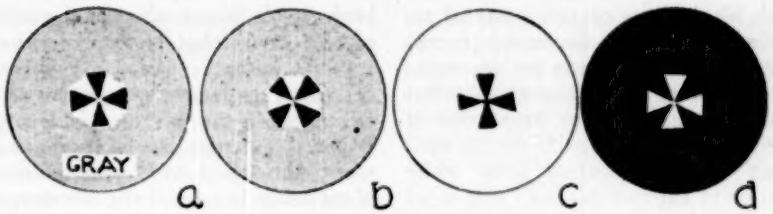


FIG. 1. The inspection and test figures

opposite 'color' (one being black on a white ground, the other white on black).

This paper reports an attempt to determine (a) whether the figure-ground reversal of an ambiguous pattern will be affected by prior satiation of the 'visual region' upon which one of the two alternative figures is to appear, and (b) whether any such effects are the same when both the inspection figure and the test region are both black, when they are both white, and when one is black and the other white.

PROCEDURE

Each subject was seated before a reduction screen permitting only monocular vision with the non-dominant eye, his head in a headrest and his left and right hands resting upon telegraph keys. A reversible figure 6.5 in. in diameter was presented against a neutral gray background which filled the field, at a distance of 32 in., and the subject was asked to press the left key while the black portion of the pattern was figure and to press the right key while the white portion of the pattern was figure. The key-pressings were recorded on a polygraph. After 10 reversals had occurred, the reduction screen was so moved that only the dominant eye was uncovered, and an inspection figure of the same size was presented, its background filling the field of vision. The subject fixated this figure until he reported certain criteria of satiation to have appeared, or until five min. had elapsed. After this, the dominant eye was covered, and the original reversible figure was exposed to the non-dominant eye, until 10 reversals had been reported by telegraph key. The subjects were divided into four groups, groups A(N = 15) and B (N = 20) having as their inspection figure the black cross, Fig. 1c, groups C (N = 20) and D (N = 17), the white cross, Fig. 1d. As reversible test patterns, groups A and C had Fig. 1a, groups B and D, Fig. 1b.

The instructions to the subject were as follows: "You will see a reversible pattern in which either the black cross is seen as figure or the white cross is seen as figure. Maintain a passive attitude and allow the figure to reverse itself as it will. Maintain a fixation point as close to the center of the figure as you can without letting the figure fall into wedges, that is, whichever cross is being seen should be seen as a complete cross. Press the left-hand telegraph key while you see the black cross and the right-hand telegraph key while you see the white cross. You may practice for a few moments until you feel that you can follow the reversals with the telegraph keys." The instructions with respect to the inspection figure were as follows: "This time you are not to press telegraph keys; you are to fixate the center of this cross, and you are to maintain an active attitude of trying to perceive the cross as figure, as a whole cross on a gray background. If, despite your attempt to the contrary, the wedges of background between the arms of the cross start becoming figure so that the pattern appears to become a field with the cross cut out of it, or any similar phenomena occur, please let me know."

RESULTS AND DISCUSSION

The relative amounts of time devoted to perception-as-figure of the white and of the black crosses (the two alternative figures within the ambiguous pattern) was obtained before and after satiation, to determine which of the two shows an increased perception-as-figure in the after-period as compared to the fore-period. (See Table I.) A significant proportion of the subjects in Group A (black satiating figure spatially coincident

with black cross of test pattern) revealed a relatively decreased perception of the black cross (or increased perception of the white cross) after satiation; a significant proportion of the subjects of Group D (white satiating figure coincident with white portion of the test pattern) evidenced a relatively decreased perception of the white cross (or increased perception of the black cross) after satiation. In Groups B (black satiation figure coincident with white cross of test pattern) and C (white satiating figure coincident with black portion of test pattern), however, the results obtained were not different from what would be expected if the satiation period had no consistent effects, since half the subjects of each group displayed an increased perception of the white cross after satiation, and half displayed an increased perception of the black cross.

Thus (as is shown in Groups A and D), if the satiating figure and the corresponding portion of the reversible pattern are of the same 'color' (i.e., both are black on a white background, or vice versa) that portion of the reversible pattern suffers a relatively decreased tendency to be perceived as figure, while (as is shown in Groups B and C) satiation has *no* effects at all when the black-white relationships of satiating figure and the corresponding portion of the test pattern are reversed. Since Köhler and Wallach hold that the only difference obtained by substituting a black figure on a white ground for a white figure on a black ground is a reversal of the direction of the figural currents involved (**2**, p. 331), and that satiation is non-directional in nature (**2**, pp. 331, 332), the results here obtained are in conflict with the process of satiation as they describe it. Several ways of resolving this difficulty suggest themselves: (a) the effects reported in this paper may have no close relation to the hypothesized satiational nature of figure-ground reversal, but may represent the working of some as-yet unknown process; (b) the figural currents arising from a black figure on

TABLE I

FREQUENCY AND PERCENTS OF SUBJECTS WHO DISPLAYED INCREASED PERCEPTION OF THE BLACK CROSS AFTER SATIATION UNDER EACH CONDITION

Group	A	B	C	D
Test-figure presented before and after satiation	1a	1b	1a	1b
Inspection figure, satiation	1c	1c	1d	1d
No. of subjects in group	20	20	20	20
No. of subjects displaying increased perception of black cross after satiation	5	11	10	16
Percent displaying increased perception of black cross after satiation	25%	55%	50.0%	80%
Difference from 50%	25%	−5%	0.0%	−30%
Standard error by null hypothesis	11.2	11.2	11.2	11.2
C.R.	2.2			2.7
P	<0.05			<0.01

white ground and those from a white figure on black ground may be different in other ways, as well as opposite in direction; (c) there may exist directional as well as non-directional components of the satiation-induced resistance. To this writer, the last hypothesis seems the most economical, since it would permit both the contour-repulsion phenomena discovered by Köhler and Wallach, and the effects of satiation upon figure-ground reversal to be described within the same system of constructs, as follows.

If the state of raised resistance occasioned by satiation-induced non-directional electrotonus were accompanied by a certain amount of something analogous to a residual polarization of the medium, this component would oppose the passage of current in the direction of the satiating current and facilitate current flow in the opposing direction. The results here obtained might then be explicable as follows: where the color (and therefore the direction of current flow) of the satiating figure and of the subsequently coincident portion of the test pattern were the same, perception-as-figure of that portion of the test figure would be opposed both by the directionless factor and the directional component; where the color (and direction of current flow) of the satiating figure and of the subsequently coincident portion of the test pattern were opposite, perception-as-figure of that portion of the test pattern would be opposed by the non-directional components and facilitated by the directional components, so that in certain situations there might be no net effect (as obtained in Groups B and C).

As evidence against any directional attributes of satiation, Köhler and Wallach have demonstrated (2, p. 332) that figural after-effects (measured at some small distance away from the figural region) caused by satiation with a black figure on a white ground are not noticeably greater than those caused by alternate presentation of a black figure on a white ground and the same figure white on a black ground (here, reversing the brightness relationship of figure and ground in the inspection figure reverses the direction of the satiating current and therefore cancels any directional components, so that the absence of any decrease in effectiveness under these circumstances is taken as indicating the non-directional nature of satiation). If, however, the directional effect were a small one compared to the non-directional component, it might only be noticed in situations depending upon as delicate a balance of factors as may be operative in figure-ground reversal (where the non-directional components of satiation must quickly become very nearly equal in the two alternative figural regions) and further, might be noticeable only in the region of greatest satiation, the figural region itself. With such characteristics, the absence of the directional effect from the situations tested by Köhler and Wallach would seem understandable.

Summary

It has been found that prior satiation of the visual region upon which one of the two alternative figures of an ambiguous test pattern is later presented (a) results in the relative decrease in perception-as-figure of the portion of the test pattern subsequently coincident with that region if both inspection figure and coincident portion of the test pattern are black on white, or both are white on black; (b) results in no change in the figure-ground reversal of the ambiguous pattern if the inspection figure is white on black and the coincident portion of test pattern is black on white, or vice versa. The findings described under (a) would seem to be in support of Köhler's suggestion that figure-ground reversals are due to a satiational process, while the absence of those effects in (b) would seem irreconcilable with a completely non-

directional satiation process as described by Köhler and Wallach. The possibility of a weak directional satiational component is suggested.

(Manuscript received for immediate publication August 9, 1950)

References

1. Köhler, W. *Dynamics in psychology.* New York: Liveright, 1940.
2. Köhler, W., & Wallach, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. philos. Soc.*, 1944, **88**, 269–357.

RULES REGARDING ANIMALS

a statement prepared by

The Committee on Precautions
in Animal Experimentation
of the
American Psychological Association

and approved by

The Council of Representatives of the
American Psychological Association

on September 8, 1949

COPIES SUITABLE FOR FRAMING

sent to

Animal Laboratories at Their Request

PSYCHOLOGICAL MONOGRAPH

#270

1945

ON PROBLEM SOLVING

By

KARL DUNCKER

$2.50

This popular monograph is #270 of the Psychological Monograph series. It has been reprinted so that it is again available.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N. W.

Washington 5, D, C.